

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OW protein - protein search, using sw model

Run on: February 28, 2002, 11:34:38 : Search time 38.63 Seconds  
(without alignments)  
36.433 Million cell updates/sec

Title: US-09-359-426C-2  
Perfect score: 81  
Sequence: 1 XEKPPLTAAAXAPVXNA 19

Scoring table: BIOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0  
Minimum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :  
1: /SIDSB/gcgdata/geneseq/geneseq/AA1980.DAT:  
2: /SIDSB/gcgdata/geneseq/geneseq/AA1981.DAT:  
3: /SIDSB/gcgdata/geneseq/geneseq/AA1982.DAT:  
4: /SIDSB/gcgdata/geneseq/geneseq/AA1983.DAT:  
5: /SIDSB/gcgdata/geneseq/geneseq/AA1984.DAT:  
6: /SIDSB/gcgdata/geneseq/geneseq/AA1985.DAT:  
7: /SIDSB/gcgdata/geneseq/geneseq/AA1986.DAT:  
8: /SIDSB/gcgdata/geneseq/geneseq/AA1987.DAT:  
9: /SIDSB/gcgdata/geneseq/geneseq/AA1988.DAT:  
10: /SIDSB/gcgdata/geneseq/geneseq/AA1989.DAT:  
11: /SIDSB/gcgdata/geneseq/geneseq/AA1990.DAT:  
12: /SIDSB/gcgdata/geneseq/geneseq/AA1991.DAT:  
13: /SIDSB/gcgdata/geneseq/geneseq/AA1992.DAT:  
14: /SIDSB/gcgdata/geneseq/geneseq/AA1993.DAT:  
15: /SIDSB/gcgdata/geneseq/geneseq/AA1994.DAT:  
16: /SIDSB/gcgdata/geneseq/geneseq/AA1995.DAT:  
17: /SIDSB/gcgdata/geneseq/geneseq/AA1996.DAT:  
18: /SIDSB/gcgdata/geneseq/geneseq/AA1997.DAT:  
19: /SIDSB/gcgdata/geneseq/geneseq/AA1998.DAT:  
20: /SIDSB/gcgdata/geneseq/geneseq/AA1999.DAT:  
21: /SIDSB/gcgdata/geneseq/geneseq/AA2000.DAT:  
22: /SIDSB/gcgdata/geneseq/geneseq/AA2001.DAT:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	76	93.8	19	AAW64480	P. aeruginosa prot
2	76	93.8	19	AAW64480	Pseudomonas aerugi
3	44	54.3	291	AAAG30091	Arabidopsis thalia
4	44	54.3	292	AAAG30090	Arabidopsis thalia
5	44	54.3	306	AAAG30089	Amino acid sequenc
6	44	54.3	427	AAAG30088	Arabidopsis thalia
7	42	51.9	172	AAAG3604	Arabidopsis thalia
8	42	51.9	215	AAAG3603	Arabidopsis thalia
9	42	51.9	332	AAAG3602	HIV-type virus thalia
10	40	49.4	498	AAW51691	HIV isolate 5180 g
11	40	49.4	20	AAW93076	

12	39	48.1	128	21	AAAG36990
13	39	48.1	233	22	AAAG1909
14	38	46.9	229	22	AAAG1775
15	38	46.9	261	22	AAAG1455
16	38	46.9	307	11	AAAR07361
17	38	46.9	521	15	AAAR63672
18	37	45.7	120	21	AAAG18092
19	37	45.7	243	21	AAAG05456
20	37	45.7	456	21	AAV54289
21	37	45.7	498	18	AAW06454
22	37	45.7	498	21	AAV34313
23	37	45.7	500	21	AAV54315
24	37	45.7	500	21	AAV54315
25	37	45.7	1064	22	AAAB46588
26	37	45.7	1064	22	AAAB46589
27	36	44.4	19	21	AAV85075
28	36	44.4	71	21	AAV85077
29	36	44.4	71	21	AAV85077
30	36	44.4	126	21	AAAB43394
31	36	44.4	128	21	AAAG22994
32	36	44.4	163	21	AAAG16982
33	36	44.4	181	17	AAW07186
34	36	44.4	182	18	AAW44795
35	36	44.4	184	21	AAV84718
36	36	44.4	187	21	AAAG16981
37	36	44.4	223	22	AAAG1790
38	36	44.4	229	20	AAV38536
39	36	44.4	250	21	AAAG24737
40	36	44.4	250	21	AAAG24737
41	36	44.4	250	21	AAAG24737
42	36	44.4	278	21	AAAG24736
43	36	44.4	278	21	AAAG24736
44	36	44.4	278	21	AAAG24736
45	36	44.4	284	21	AAAG36899

## ALIGNMENTS

RESULT 1	
ID	AAW64480 standard; peptide: 19 AA.
XX	AAW64480:
AC	20-OCN-1998 (first entry)
XX	P. aeruginosa protein antigen Pa60 N-terminal peptide fragment.
XX	Antigen: Pa60: diagnosis; detection: cystic fibrosis; vaccine.
KW	Immunogen: Infection; treatment.
XX	Pseudomonas aeruginosa.
OS	
XX	
XX	Key
XX	Misc-difference 1 Location/Qualifiers
FT	Misc-difference 1 /label- unknown
FT	Misc-difference 12 /label- unknown
FT	Misc-difference 17 /label- unknown
FT	Misc-difference 17 /label- unknown
XX	
PN	WO9832769-A1.
XX	30-JUL-1998.
PD	26-JAN-1998: 98WO-GB00217.
XX	
PE	24-JAN-1997: 97GB-0001489.
XX	
PR	(AUSP-) AUSPHARM INT LTD.
XX	(CHAP/) CHAPMAN P W.
PA	
XX	

Arabidopsis thalia  
C glutamic prote  
Human protein sequ  
C glutamic prote  
Phospholipase D to  
Aldehyde-dehydroge  
Arabidopsis thalia  
Consensus sequence  
Capsicum annum 1y  
Amino acid sequenc  
Amino acid sequenc  
Serine protease in  
Immunogenic peptid  
FMV non-structura  
Human ORF3158  
Arabidopsis thalia  
Arabidopsis thalia  
B. atzei1 strain  
A. decolin binding  
Arabidopsis thalia  
S. epidermidis ope  
Neisseria gonorrhoe  
Arabidopsis thalia  
Arabidopsis thalia  
Arabidopsis thalia  
Arabidopsis thalia  
Arabidopsis thalia  
Arabidopsis thalia

PI Clancy RL, Cripps AM, Dunkley M, Kyd J;  
 DR WPI; 1998-427879/36.  
 XX  
 XX Protein antigen from *Pseudomonas aeruginosa* and its antigenic  
 PT fragments: useful diagnostically to detect specific antibodies,  
 PR particularly in patients with cystic fibrosis, and as vaccines  
 XX  
 PS Claim 3; Page 2; 23pp; English.

CC This peptide is the N-terminal fragment of a novel *Pseudomonas aeruginosa*  
 CC protein antigen, Pa60. This fragment could be used for diagnostic  
 CC detection of *P. aeruginosa* by forming complexes with specific antibodies,  
 CC particularly in patients with cystic fibrosis (by analysis of mucus, e.g.  
 CC in saliva), or in vaccines or immunogenic compositions to treat or  
 CC prevent infection by *P. aeruginosa*.

Sequence 19 AA;

Query Match 93.8%; Score 76; DB 19; Length 19;  
 Best Local Similarity 100.0%; Pred. No. 1,7e-06;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 EKKTPLTAAAPVXNA 19  
 |||||  
 DB 2 eektplttaaxapvxna 19

RESULT 2

AAB69062 standard; peptide; 19 AA.

AC AAB69062;

DT 18-APR-2001 (first entry)

DE *Pseudomonas aeruginosa* protein N-terminal peptide.

KW *Pseudomonas aeruginosa*; chitinase; groEL; chIA; antigen; vaccine;  
 KM diagnosis; detection; infection; immune response.

OS *Pseudomonas aeruginosa*.

Key Location/Qualifiers

FT Misc-difference 12 /note= "unspecified"

FT Misc-difference 17 /note= "unspecified"

FT Misc-difference 17 /note= "unspecified"

PN WO200102577-A1.

PD 11-JAN-2001.

PF 03-JUL-2000; 2000MO-GB02554.

PR 01-JUL-1999; 99GB-0015419.

PA (PROV-) PROVALIS UK LTD.

PI Smith CJ, Thompson SE, Smith MW, Peek K, Sizer PJH, Wilkinson MC;  
 DR WPI; 2001-080988/09.

PT Antigenic *Pseudomonas aeruginosa* proteins, useful in the detection  
 PT and/or diagnosis of *P. aeruginosa* infections and for producing vaccines  
 PR against *P. aeruginosa*.  
 XX Disclosure; Page 2; 129pp; English.

\*CC The present invention describes antigenic *Pseudomonas aeruginosa*

CC proteins (P1). The *P. aeruginosa* proteins have antibacterial activity  
 CC and can be used in vaccines and as antagonists. The proteins or their  
 CC fragments, or antibodies are useful in the detection and/or diagnosis  
 CC of *P. aeruginosa*. They are also useful for producing a vaccine and  
 CC inducing an immune response against *P. aeruginosa* infection. An agent  
 CC capable of antagonising, inhibiting or otherwise interfering with the  
 CC function or expression of P1 are useful in the manufacture of a  
 CC medicament for the treatment or prophylaxis of *P. aeruginosa* infections.  
 CC The present sequence represents a probable *P. aeruginosa* protein  
 CC N-terminal peptide sequence from the present invention.

Sequence 19 AA;

Query Match 93.8%; Score 76; DB 22; Length 19;  
 Best Local Similarity 100.0%; Pred. No. 1,7e-06;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 EKKTPLTAAAPVXNA 19  
 |||||  
 DB 2 eektplttaaxapvxna 19

RESULT 3

AAG30091 standard; Protein; 291 AA.

AC AAG30091;

DT 17-OCT-2000 (first entry)

DE *Arabidopsis thaliana* protein fragment SEQ ID NO: 35912.

KW Protein identification; signal transduction pathway; metabolic pathway;  
 KM hybridisation assay; genetic mapping; gene expression control; promoter;  
 XX termination sequence.

OS *Arabidopsis thaliana*.

PN EPI033405-A2.

PD 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439.

PR 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130510.

PR 23-APR-1999; 99US-0130891.

PR 28-APR-1999; 99US-0131449.

PR 30-APR-1999; 99US-0132048.

PR 04-MAY-1999; 99US-0132407.

PR 05-MAY-1999; 99US-0132484.

PR 06-MAY-1999; 99US-0132485.

PR 06-MAY-1999; 99US-0132486.

PR 07-MAY-1999; 99US-0132487.

PR 11-MAY-1999; 99US-0132863.

PR 14-MAY-1999; 99US-0134226.

PR 14-MAY-1999; 99US-0134218.

PR 14-MAY-1999; 99US-0134219.

PR 14-MAY-1999; 99US-0134221.

PR 14-MAY-1999; 99US-0134370.

PR 18-MAY-1999; 99US-0134768.

PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136382.  
PR 28-MAY-1999; 99US-0137222.  
PR 01-JUN-1999; 99US-0137528.  
PR 03-JUN-1999; 99US-0137502.  
PR 04-JUN-1999; 99US-0137724.  
PR 07-JUN-1999; 99US-0138094.  
PR 08-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 18-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140685.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140981.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 19-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 21-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 22-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.

PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148585.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 27-AUG-1999; 99US-0151081.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151530.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159285.  
PR 13-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 18-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 21-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 22-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.

PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161920.  
PR 29-OCT-1999; 99US-0162142.

Query Match 54.3% Score 44; DB 21; Length 291;  
Best Local Similarity 60.0%; Pred. No. 9.7;  
Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 2 EKKTLTTRAAAPV 16  
Db 72 eekamtlamspv 86

PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161920.  
PR 29-OCT-1999; 99US-0162142.

17-OCT-2000 (first entry)  
Arabidopsis thaliana protein fragment SEQ ID NO: 35911.

Protein identification: signal transduction pathway; metabolic pathway;  
hybridisation assay; genetic mapping; gene expression control; promoter;  
termination sequence.

Arabidopsis thaliana.

EP1033405-A2.

06-SEP-2000.

25-FEB-2000; 2000EP-0301439.

25-FEB-1999; 99US-0121825.

05-MAR-1999; 99US-0123180.

09-MAR-1999; 99US-0123548.

23-MAR-1999; 99US-0125768.

25-MAR-1999; 99US-0126264.

29-MAR-1999; 99US-0126785.

01-APR-1999; 99US-0127462.

08-APR-1999; 99US-0128234.

16-APR-1999; 99US-0128845.

19-APR-1999; 99US-0130077.

21-APR-1999; 99US-0130449.

23-APR-1999; 99US-0130891.

28-APR-1999; 99US-0131449.

30-APR-1999; 99US-0132048.

04-MAY-1999; 99US-0132407.

05-MAY-1999; 99US-0132484.

06-MAY-1999; 99US-0132485.

07-MAY-1999; 99US-0132486.

11-MAY-1999; 99US-0132863.

14-MAY-1999; 99US-0134256.

14-MAY-1999; 99US-0134256.

14-MAY-1999; 99US-0134256.

14-MAY-1999; 99US-0134256.

14-MAY-1999; 99US-0134256.

PR 01-JUN-1999; 99US-0137222.

PR 03-JUN-1999; 99US-0137528.

PR 04-JUN-1999; 99US-0137502.

PR 07-JUN-1999; 99US-0137724.

PR 08-JUN-1999; 99US-0138094.

PR 10-JUN-1999; 99US-0138540.

PR 10-JUN-1999; 99US-0138847.

PR 14-JUN-1999; 99US-0139119.

PR 16-JUN-1999; 99US-0139452.

PR 17-JUN-1999; 99US-0139453.

PR 18-JUN-1999; 99US-0139454.

PR 18-JUN-1999; 99US-0139455.

PR 18-JUN-1999; 99US-0139456.

PR 18-JUN-1999; 99US-0139457.

PR 18-JUN-1999; 99US-0139458.

PR 18-JUN-1999; 99US-0139459.

PR 18-JUN-1999; 99US-0139460.

PR 18-JUN-1999; 99US-0139461.

PR 18-JUN-1999; 99US-0139462.

PR 18-JUN-1999; 99US-0139463.

PR 18-JUN-1999; 99US-0139464.

PR 18-JUN-1999; 99US-0139465.

PR 18-JUN-1999; 99US-0139466.

PR 18-JUN-1999; 99US-0139467.

PR 18-JUN-1999; 99US-0139468.

PR 18-JUN-1999; 99US-0139469.

PR 18-JUN-1999; 99US-0139470.

PR 18-JUN-1999; 99US-0139471.

PR 18-JUN-1999; 99US-0139472.

PR 18-JUN-1999; 99US-0139473.

PR 18-JUN-1999; 99US-0139474.

PR 18-JUN-1999; 99US-0139475.

PR 18-JUN-1999; 99US-0139476.

PR 18-JUN-1999; 99US-0139477.

PR 18-JUN-1999; 99US-0139478.

PR 18-JUN-1999; 99US-0139479.

PR 18-JUN-1999; 99US-0139480.

PR 18-JUN-1999; 99US-0139481.

PR 18-JUN-1999; 99US-0139482.

PR 18-JUN-1999; 99US-0139483.

PR 18-JUN-1999; 99US-0139484.

PR 18-JUN-1999; 99US-0139485.

PR 18-JUN-1999; 99US-0139486.

PR 18-JUN-1999; 99US-0139487.



PR 05-AUG-1999; 99US-0147192.  
 PR 05-AUG-1999; 99US-0147260.  
 PR 06-AUG-1999; 99US-0147303.  
 PR 06-AUG-1999; 99US-0147416.  
 PR 09-AUG-1999; 99US-0147493.  
 PR 09-AUG-1999; 99US-0147935.  
 PR 10-AUG-1999; 99US-0148171.  
 PR 11-AUG-1999; 99US-0148319.  
 PR 12-AUG-1999; 99US-0148341.  
 PR 13-AUG-1999; 99US-0148565.  
 PR 13-AUG-1999; 99US-0148684.  
 PR 16-AUG-1999; 99US-0149368.  
 PR 17-AUG-1999; 99US-0149175.  
 PR 18-AUG-1999; 99US-0149426.  
 PR 20-AUG-1999; 99US-0149722.  
 PR 20-AUG-1999; 99US-0149723.  
 PR 20-AUG-1999; 99US-0149929.  
 PR 23-AUG-1999; 99US-0149902.  
 PR 23-AUG-1999; 99US-0149930.  
 PR 25-AUG-1999; 99US-0150566.  
 PR 26-AUG-1999; 99US-0150884.  
 PR 27-AUG-1999; 99US-0151065.  
 PR 27-AUG-1999; 99US-0151066.  
 PR 27-AUG-1999; 99US-0151080.  
 PR 30-AUG-1999; 99US-0151303.  
 PR 31-AUG-1999; 99US-0151930.  
 PR 01-SEP-1999; 99US-0152353.  
 PR 07-SEP-1999; 99US-0153070.  
 PR 10-SEP-1999; 99US-0153070.  
 PR 13-SEP-1999; 99US-0153758.  
 PR 15-SEP-1999; 99US-0154018.  
 PR 16-SEP-1999; 99US-0154039.  
 PR 16-SEP-1999; 99US-0154779.  
 PR 20-SEP-1999; 99US-0155139.  
 PR 22-SEP-1999; 99US-0155139.  
 PR 23-SEP-1999; 99US-0155486.  
 PR 24-SEP-1999; 99US-0155659.  
 PR 28-SEP-1999; 99US-0156458.  
 PR 29-SEP-1999; 99US-0156596.  
 PR 04-OCT-1999; 99US-0157117.  
 PR 05-OCT-1999; 99US-0157753.  
 PR 06-OCT-1999; 99US-0157865.  
 PR 07-OCT-1999; 99US-0158029.  
 PR 08-OCT-1999; 99US-0158232.  
 PR 12-OCT-1999; 99US-0158369.  
 PR 13-OCT-1999; 99US-0159293.  
 PR 13-OCT-1999; 99US-0159294.  
 PR 13-OCT-1999; 99US-0159295.  
 PR 14-OCT-1999; 99US-0159329.  
 PR 14-OCT-1999; 99US-0159330.  
 PR 14-OCT-1999; 99US-0159331.  
 PR 14-OCT-1999; 99US-0159637.  
 PR 14-OCT-1999; 99US-0159638.  
 PR 18-OCT-1999; 99US-0159584.  
 PR 21-OCT-1999; 99US-0160741.  
 PR 21-OCT-1999; 99US-0160767.  
 PR 21-OCT-1999; 99US-0160768.  
 PR 21-OCT-1999; 99US-0160770.  
 PR 21-OCT-1999; 99US-0160814.  
 PR 21-OCT-1999; 99US-0160815.  
 PR 21-OCT-1999; 99US-0160815.  
 PR 22-OCT-1999; 99US-0160980.  
 PR 22-OCT-1999; 99US-0160981.  
 PR 22-OCT-1999; 99US-0160989.  
 PR 22-OCT-1999; 99US-0161404.  
 PR 23-OCT-1999; 99US-0161405.  
 PR 23-OCT-1999; 99US-0161406.  
 PR 26-OCT-1999; 99US-0161359.  
 PR 26-OCT-1999; 99US-0161360.  
 PR 26-OCT-1999; 99US-0161361.  
 PR 28-OCT-1999; 99US-0161920.  
 PR 28-OCT-1999; 99US-0161992.  
 PR 28-OCT-1999; 99US-0161993.  
 PR 29-OCT-1999; 99US-0162142.

Query Match 54.3%; Score 44; DB 21; Length 292;  
 Best Local Similarity 60.0%; Pred. No. 9.7;  
 Matches 9; Conservative 2; Mismatches 4; Indels 0; Caps 0;

QY 2 EKKPLTTAAXAPV 16  
 ||| :||| :|||

Db 73 eekaemtcmqspv 87

# RESULT 5

AA67588  
 ID AAB67588 standard; Protein; 306 AA.

AC AAB67588;

DT 29-MAY-2001 (first entry)

XX Amino acid sequence of a deoxyribokinase enzyme.

XX Deoxyribonucleoside: deoxyribose 1-phosphate; thymidine phosphorylase;

KW putine nucleoside phosphorylase; phosphopentose mutase;

KW phosphopentose aldolase; fructose 1,6-diphosphate aldolase;

KW deoxyribokinase; nucleoside 2-deoxyribosyltransferase.

XX Salmonella typhi.

OS WO200114566-A2.

PN 01-MAR-2001.

PD 18-AUG-2000; 2000MO-EP08088.

PE 20-AUG-1999; 99EP-0116425.

XX (HOPE ) ROCHE DIAGNOSTICS GMBH.

PA (INSP ) INST PASTEUR.

PA (PHAR ) PHARMA-WALDHOF GMBH & CO KG.

XX Tischer W, Ihlenfeldt H, Barzu O, Sakamoto H, Pistotnik E;

PI Marliere P, Pochet S;

XX WPI: 2001-235026/24.

DR N-PSDB; AAF5444.

XX In vitro enzymatic synthesis of deoxyribonucleosides comprises reacting

PT deoxyribose 1-phosphate and a nucleobase to form a deoxyribonucleoside

PT and an inorganic phosphate

XX Disclosure: Page 59-61; 73pp; English.

PS The present sequence represents a deoxyribokinase enzyme. This enzyme

XX is involved in the biosynthesis of deoxyribonucleosides, and is

CC used in the method of the invention. The specification describes a

CC method for the in vitro enzymatic synthesis of deoxyribonucleosides

CC The method comprises reacting deoxyribose 1-phosphate and a nucleobase

CC to form a deoxyribonucleoside and an inorganic phosphate. Enzymes which

CC may be used in the method of the invention include thymidine

CC phosphorylase, putine nucleoside phosphorylase, phosphopentose mutase,

CC phosphopentose aldolase, fructose 1,6-diphosphate aldolase,

CC deoxyribokinase, and nucleoside 2-deoxyribosyltransferase.

XX Sequence 306 AA;

SQ

Query Match 54.3%; Score 44; DB 22; Length 306;  
 Best Local Similarity 47.1%; Pred. No. 10;

Matches 8; Conservative 3; Mismatches 6; Indels 0; Caps 0;

QY 3 EKKPLTTAAXAPVYXNA 19  
 ||| :||| :|||

Db 84 ekvpctssvafivfna 100

RESULT 6  
AAG30089 standard; Protein; 427 AA.  
AC AAG30089;  
XX  
XX 17-OCT-2000 (first entry)  
DE Arabidopsis thaliana protein fragment SFQ ID NO: 35910.  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
XX termination sequence.  
OS Arabidopsis thaliana.  
XX  
XX EPI033405-A2.  
XX  
XX 06-SEP-2000.  
XX  
XX 25-FEB-2000; 2000EP-0301439.  
XX  
XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125768.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130049.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 04-MAY-1999; 99US-0132407.  
PR 05-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 11-MAY-1999; 99US-0132863.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 20-MAY-1999; 99US-0134941.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 02-JUL-1999; 99US-0142154.  
PR 06-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 20-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 21-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0145086.  
PR 22-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 26-JUL-1999; 99US-0145224.  
PR 27-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 28-JUL-1999; 99US-0145919.  
PR 02-AUG-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 03-AUG-1999; 99US-0146389.  
PR 04-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 05-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147312.  
PR 06-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.

PR 20-AUG-1999: 99US-0149929.  
PR 23-AUG-1999: 99US-0149930.  
PR 23-AUG-1999: 99US-0149930.  
PR 25-AUG-1999: 99US-0150566.  
PR 26-AUG-1999: 99US-0150884.  
PR 27-AUG-1999: 99US-0151065.  
PR 27-AUG-1999: 99US-0151066.  
PR 27-AUG-1999: 99US-0151080.  
PR 30-AUG-1999: 99US-0151303.  
PR 31-AUG-1999: 99US-0151338.  
PR 01-SEP-1999: 99US-0151930.  
PR 07-SEP-1999: 99US-0152363.  
PR 10-SEP-1999: 99US-0153070.  
PR 13-SEP-1999: 99US-0153758.  
PR 15-SEP-1999: 99US-0154018.  
PR 16-SEP-1999: 99US-0154039.  
PR 20-SEP-1999: 99US-0154779.  
PR 22-SEP-1999: 99US-0155139.  
PR 23-SEP-1999: 99US-0155486.  
PR 24-SEP-1999: 99US-0155659.  
PR 28-SEP-1999: 99US-0156458.  
PR 29-SEP-1999: 99US-0156596.  
PR 04-OCT-1999: 99US-0157117.  
PR 05-OCT-1999: 99US-0157753.  
PR 06-OCT-1999: 99US-0157865.  
PR 07-OCT-1999: 99US-0158029.  
PR 08-OCT-1999: 99US-0158232.  
PR 12-OCT-1999: 99US-0158369.  
PR 13-OCT-1999: 99US-0159283.  
PR 13-OCT-1999: 99US-0159294.  
PR 13-OCT-1999: 99US-0159295.  
PR 14-OCT-1999: 99US-0159329.  
PR 14-OCT-1999: 99US-0159330.  
PR 14-OCT-1999: 99US-0159331.  
PR 14-OCT-1999: 99US-0159637.  
PR 14-OCT-1999: 99US-0159638.  
PR 18-OCT-1999: 99US-0159584.  
PR 21-OCT-1999: 99US-0160741.  
PR 21-OCT-1999: 99US-0160767.  
PR 21-OCT-1999: 99US-0160768.  
PR 21-OCT-1999: 99US-0160770.  
PR 21-OCT-1999: 99US-0160814.  
PR 21-OCT-1999: 99US-0160815.  
PR 21-OCT-1999: 99US-0160815.  
PR 22-OCT-1999: 99US-0160981.  
PR 22-OCT-1999: 99US-0160989.  
PR 22-OCT-1999: 99US-0161404.  
PR 25-OCT-1999: 99US-0161405.  
PR 25-OCT-1999: 99US-0161406.  
PR 26-OCT-1999: 99US-0161359.  
PR 26-OCT-1999: 99US-0161360.  
PR 26-OCT-1999: 99US-0161361.  
PR 28-OCT-1999: 99US-0161920.  
PR 28-OCT-1999: 99US-0161922.  
PR 28-OCT-1999: 99US-0161993.  
PR 29-OCT-1999: 99US-0162142.

Query Match 54.38: Score 44: DB 21: Length 427:  
Best Local Similarly 60.08: Pred. No. 15:  
Matches 9: Conservative 2: Mismatches 4: Indels 0: Gaps 0:

OY 2 EKKTPITTAAXAPV 16  
DB 208 eekamtlamqspv 222

RESULT 7  
ID AAG43604 standard: Protein: 172 AA.  
XX AAG43604:  
AC  
XX  
DT 18-OCT-2000 (first entry)

XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 54519.  
XX Protein identification; signal transduction pathway; metabolic pathway;  
XX hybridisation assay; genetic mapping; gene expression control; promoter;  
XX termination sequence.  
OS Arabidopsis thaliana.  
XX EP1033405-A2.  
XX PD 06-SEP-2000.  
XX PF 25-FEB-2000: 2000EP-0301439.  
XX 25-FEB-1999: 99US-0121825.  
XX 05-MAR-1999: 99US-0123180.  
XX 09-MAR-1999: 99US-0123548.  
XX 23-MAR-1999: 99US-0125788.  
XX 25-MAR-1999: 99US-0126264.  
XX 29-MAR-1999: 99US-0126785.  
XX 01-APR-1999: 99US-0127462.  
XX 06-APR-1999: 99US-0128234.  
XX 08-APR-1999: 99US-0128714.  
XX 16-APR-1999: 99US-0129845.  
XX 19-APR-1999: 99US-0130077.  
XX 21-APR-1999: 99US-0130449.  
XX 23-APR-1999: 99US-0130510.  
XX 28-APR-1999: 99US-0131449.  
XX 30-APR-1999: 99US-0132048.  
XX 30-APR-1999: 99US-0132407.  
XX 04-MAY-1999: 99US-0132485.  
XX 05-MAY-1999: 99US-0132485.  
XX 06-MAY-1999: 99US-0132486.  
XX 06-MAY-1999: 99US-0132487.  
XX 07-MAY-1999: 99US-0132863.  
XX 11-MAY-1999: 99US-0134256.  
XX 14-MAY-1999: 99US-0134218.  
XX 14-MAY-1999: 99US-0134221.  
XX 14-MAY-1999: 99US-0134370.  
XX 18-MAY-1999: 99US-0134768.  
XX 19-MAY-1999: 99US-0134941.  
XX 20-MAY-1999: 99US-0135353.  
XX 21-MAY-1999: 99US-0135629.  
XX 24-MAY-1999: 99US-0136021.  
XX 25-MAY-1999: 99US-0136392.  
XX 27-MAY-1999: 99US-0136782.  
XX 28-MAY-1999: 99US-0137222.  
XX 01-JUN-1999: 99US-0137528.  
XX 03-JUN-1999: 99US-0137602.  
XX 04-JUN-1999: 99US-0137724.  
XX 07-JUN-1999: 99US-0138094.  
XX 08-JUN-1999: 99US-0138540.  
XX 10-JUN-1999: 99US-0138847.  
XX 10-JUN-1999: 99US-0139119.  
XX 14-JUN-1999: 99US-0139452.  
XX 16-JUN-1999: 99US-0139453.  
XX 17-JUN-1999: 99US-0139492.  
XX 18-JUN-1999: 99US-0139454.  
XX 18-JUN-1999: 99US-0139455.  
XX 18-JUN-1999: 99US-0139456.  
XX 18-JUN-1999: 99US-0139457.  
XX 18-JUN-1999: 99US-0139458.  
XX 18-JUN-1999: 99US-0139459.  
XX 18-JUN-1999: 99US-0139460.  
XX 18-JUN-1999: 99US-0139461.  
XX 18-JUN-1999: 99US-0139462.  
XX 18-JUN-1999: 99US-0139463.  
XX 18-JUN-1999: 99US-0139750.  
XX 18-JUN-1999: 99US-0139763.

PR	27-AUG-1999;	99US-0151080.
PR	30-AUG-1999;	99US-0151303.
PR	31-AUG-1999;	99US-0151438.
PR	01-SEP-1999;	99US-0151930.
PR	07-SEP-1999;	99US-0152363.
PR	10-SEP-1999;	99US-0153070.
PR	13-SEP-1999;	99US-0153758.
PR	15-SEP-1999;	99US-0154018.
PR	16-SEP-1999;	99US-0154039.
PR	20-SEP-1999;	99US-0154779.
PR	22-SEP-1999;	99US-0155139.
PR	23-SEP-1999;	99US-0155486.
PR	24-SEP-1999;	99US-0155659.
PR	28-SEP-1999;	99US-0156458.
PR	29-SEP-1999;	99US-0156596.
PR	04-OCT-1999;	99US-0157117.
PR	05-OCT-1999;	99US-0157753.
PR	06-OCT-1999;	99US-0158029.
PR	07-OCT-1999;	99US-0158232.
PR	08-OCT-1999;	99US-0158366.
PR	12-OCT-1999;	99US-0158293.
PR	13-OCT-1999;	99US-0158294.
PR	13-OCT-1999;	99US-0159295.
PR	14-OCT-1999;	99US-0159330.
PR	14-OCT-1999;	99US-0159331.
PR	14-OCT-1999;	99US-0159637.
PR	14-OCT-1999;	99US-0159638.
PR	18-OCT-1999;	99US-0159584.
PR	21-OCT-1999;	99US-0160741.
PR	21-OCT-1999;	99US-0160767.
PR	21-OCT-1999;	99US-0160768.
PR	21-OCT-1999;	99US-0160770.
PR	21-OCT-1999;	99US-0160814.
PR	21-OCT-1999;	99US-0160815.
PR	22-OCT-1999;	99US-0160980.
PR	22-OCT-1999;	99US-0160981.
PR	22-OCT-1999;	99US-0160989.
PR	25-OCT-1999;	99US-0161404.
PR	25-OCT-1999;	99US-0161405.
PR	25-OCT-1999;	99US-0161406.
PR	26-OCT-1999;	99US-0161359.
PR	26-OCT-1999;	99US-0161360.
PR	26-OCT-1999;	99US-0161361.
PR	28-OCT-1999;	99US-0161920.
PR	28-OCT-1999;	99US-0161922.
PR	28-OCT-1999;	99US-0161993.
PR	29-OCT-1999;	99US-0162142.

Query Match	51.98;	Score 42;	DB 21;	Length 172;
Best Local Similarity	50.08;	Pred. No. 12;		
Matches 9;	Conservative 1;	Mismatches 8;	Indels 0;	Gaps 0;
QY	2	EERTPLTTAAAFVYVXNA 19		
		:		
Db	131	eeektyveektgyvpyvkkka 148		
RESULT#	8			
AAG43603				
ID	AAG43603	standard; Protein; 215 AA.		
XX				
AC	AAG43603;			
XX				
DF	18-OCT-2000	(first entry)		
XX				
DE	Arabidopsis thaliana protein fragment SEQ ID NO: 54518.			
XX				
KW	Protein identification; signal transduction pathway; metabolic pathway;			
KW	hybridisation assay; genetic mapping; gene expression control; promoter;			
XX	termination control;			
XX				

[illegible]

PR	30-JUN-1999	99US-011827
PR	01-JUL-1999	99US-014284
PR	01-JUL-1999	99US-014254
PR	02-JUL-1999	99US-014205
PR	06-JUL-1999	99US-0142390
PR	08-JUL-1999	99US-0142803
PR	09-JUL-1999	99US-0142920
PR	13-JUL-1999	99US-0142977
PR	14-JUL-1999	99US-0143542
PR	15-JUL-1999	99US-0143624
PR	16-JUL-1999	99US-0144005
PR	16-JUL-1999	99US-0144085
PR	16-JUL-1999	99US-0144086
PR	16-JUL-1999	99US-0144335
PR	19-JUL-1999	99US-0144331
PR	19-JUL-1999	99US-0144332
PR	19-JUL-1999	99US-0144333
PR	19-JUL-1999	99US-0144334
PR	19-JUL-1999	99US-0144335
PR	19-JUL-1999	99US-0144332
PR	20-JUL-1999	99US-0144632
PR	20-JUL-1999	99US-0144814
PR	21-JUL-1999	99US-0144884
PR	21-JUL-1999	99US-0145086
PR	21-JUL-1999	99US-0145088
PR	22-JUL-1999	99US-0145085
PR	22-JUL-1999	99US-0145087
PR	22-JUL-1999	99US-0145089
PR	22-JUL-1999	99US-0145129
PR	22-JUL-1999	99US-0145145
PR	22-JUL-1999	99US-0145214
PR	23-JUL-1999	99US-0145224
PR	23-JUL-1999	99US-0145276
PR	26-JUL-1999	99US-0145513
PR	27-JUL-1999	99US-0145518
PR	27-JUL-1999	99US-0145519
PR	27-JUL-1999	99US-0145551
PR	28-JUL-1999	99US-0145386
PR	02-AUG-1999	99US-0146388
PR	02-AUG-1999	99US-0146389
PR	03-AUG-1999	99US-0147038
PR	04-AUG-1999	99US-0147204
PR	04-AUG-1999	99US-0147302
PR	05-AUG-1999	99US-0147192
PR	05-AUG-1999	99US-0147260
PR	06-AUG-1999	99US-0147416
PR	06-AUG-1999	99US-0147433
PR	09-AUG-1999	99US-0147935
PR	10-AUG-1999	99US-0148111
PR	11-AUG-1999	99US-0148319
PR	12-AUG-1999	99US-0148341
PR	13-AUG-1999	99US-0148555
PR	13-AUG-1999	99US-0148684
PR	16-AUG-1999	99US-0148366
PR	17-AUG-1999	99US-0148175
PR	18-AUG-1999	99US-0148426
PR	20-AUG-1999	99US-0149372
PR	20-AUG-1999	99US-0149373
PR	20-AUG-1999	99US-0149929
PR	23-AUG-1999	99US-0149902
PR	23-AUG-1999	99US-0149930
PR	26-AUG-1999	99US-0150566
PR	26-AUG-1999	99US-0150884
PR	27-AUG-1999	99US-0151065
PR	27-AUG-1999	99US-0151066
PR	27-AUG-1999	99US-0151083
PR	27-AUG-1999	99US-0151103
PR	31-AUG-1999	99US-0151338
PR	01-SEP-1999	99US-0151930
PR	07-SEP-1999	99US-0152263
PR	10-SEP-1999	99US-0153070
PR	13-SEP-1999	99US-0153758

Query Match	51.98;	Score 42;	DB 21;	Length 215;
Best Local Similarity	50.0%;	Pred. No. 15;		
Matches	9;	Conservative	1;	Mismatches
			8;	Indels
				0;
Gaps				0;
0y	2	EERTPLTAAKAVVANA	19	
Db	174	eektpeektgyvvvka	191	

RESULT	9	
AAAG43602		
ID	AAAG43602	standard; Protein; 332 AA.
XX		
AC	AAAG43602;	
XX		
DT	18-OCT-2000	(first entry)
XX		
DE	Arabidopsis thaliana protein fragment SEQ ID NO: 54517.	
XX		
KM	Protein identification; signal transduction pathway; metabolic pathway;	
KM	hybridisation assay; genetic mapping; gene expression control; promoter;	
KW	termination sequence.	
XX		
OS	Arabidopsis thaliana.	
XX		
PN	EP1033405-A2.	
XX		
PD	06-SEP-2000.	
XX		
PF	25-FEB-2000; 2000EP-0301439.	

XX	25-FEB-1999	9905-0121625
PR	05-MAR-1999	9905-0123180
PR	09-MAR-1999	9905-0123548
PR	23-MAR-1999	9905-0125748
PR	25-MAR-1999	9905-0126654
PR	29-MAR-1999	9905-0126785
PR	01-APR-1999	9905-0127652
PR	06-APR-1999	9905-0128354
PR	08-APR-1999	9905-0129845
PR	16-APR-1999	9905-0130777
PR	23-APR-1999	9905-0130449
PR	23-APR-1999	9905-0130510
PR	28-APR-1999	9905-0130891
PR	30-APR-1999	9905-0131444
PR	04-MAY-1999	9905-0132404
PR	05-MAY-1999	9905-0132484
PR	06-MAY-1999	9905-0132485
PR	06-MAY-1999	9905-0132486
PR	07-MAY-1999	9905-0132487
PR	11-MAY-1999	9905-0132863
PR	14-MAY-1999	9905-0134256
PR	14-MAY-1999	9905-0134218
PR	14-MAY-1999	9905-0134221
PR	18-MAY-1999	9905-0134077
PR	19-MAY-1999	9905-0134370
PR	20-MAY-1999	9905-0134768
PR	21-MAY-1999	9905-0134941
PR	24-MAY-1999	9905-0135324
PR	25-MAY-1999	9905-0135353
PR	25-MAY-1999	9905-0135629
PR	27-MAY-1999	9905-0136021
PR	28-MAY-1999	9905-0136392
PR	03-JUN-1999	9905-0136782
PR	03-JUN-1999	9905-0137222
PR	07-JUN-1999	9905-0137528
PR	08-JUN-1999	9905-0137724
PR	10-JUN-1999	9905-0138094
PR	10-JUN-1999	9905-0138540
PR	14-JUN-1999	9905-0138847
PR	16-JUN-1999	9905-0139319
PR	17-JUN-1999	9905-0139452
PR	18-JUN-1999	9905-0138455
PR	18-JUN-1999	9905-0139452
PR	18-JUN-1999	9905-0139454
PR	18-JUN-1999	9905-0139455
PR	18-JUN-1999	9905-0139456
PR	18-JUN-1999	9905-0139457
PR	18-JUN-1999	9905-0139458
PR	18-JUN-1999	9905-0139459
PR	18-JUN-1999	9905-0139460
PR	18-JUN-1999	9905-0139461
PR	18-JUN-1999	9905-0139462
PR	18-JUN-1999	9905-0139463
PR	18-JUN-1999	9905-0139464
PR	22-JUN-1999	9905-0139689
PR	23-JUN-1999	9905-0140353
PR	23-JUN-1999	9905-0140354
PR	24-JUN-1999	9905-0140695
PR	29-JUN-1999	9905-0140923
PR	30-JUN-1999	9905-0141287
PR	01-JUL-1999	9905-0141842
PR	01-JUL-1999	9905-0142154
PR	02-JUL-1999	9905-0142055
PR	06-JUL-1999	9905-0142390
PR	08-JUL-1999	9905-0142803
PR	09-JUL-1999	9905-0142920

```

PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 17-JUL-1999; 99US-0144086.
PR 18-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 19-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 21-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 23-JUL-1999; 99US-0145276.
PR 26-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145919.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 20-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 23-AUG-1999; 99US-0150566.
PR 25-AUG-1999; 99US-0150884.
PR 26-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 27-AUG-1999; 99US-0151303.
PR 30-AUG-1999; 99US-0151438.
PR 31-AUG-1999; 99US-0151930.
PR 01-SEP-1999; 99US-0152353.
PR 07-SEP-1999; 99US-0153070.
PR 10-SEP-1999; 99US-0153078.
PR 13-SEP-1999; 99US-0154018.
PR 15-SEP-1999; 99US-0154039.
PR 16-SEP-1999; 99US-0154779.
PR 20-SEP-1999; 99US-0155139.
PR 22-SEP-1999; 99US-0155486.
PR 23-SEP-1999; 99US-0155659.
PR 24-SEP-1999; 99US-0156458.
PR 28-SEP-1999;

```

```

PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 14-OCT-1999; 99US-0159584.
PR 18-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 21-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161359.
PR 25-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 26-OCT-1999; 99US-0161361.
PR 26-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161993.
PR 28-OCT-1999; 99US-0162142.
PR 29-OCT-1999;

```

```

Query Match 51.9%; Score 42; DB 21; Length 332;
Best Local Similarity 50.0%; Pred. No. 25;
Matches 9; Conservative 1; Mismatches 8; Indels 0; Caps 0;

```

```

OY 2 EEEKPLTTAAKAPVXNA 19
| | | | | | | | | |
Db 291 eektpvekktygvvka 308

```

```

RESULT 10
ID AAR51691 standard; Protein: 498 AA.
XX
AC AAR51691:
XX
DT 07-NOV-1994 (first entry)
XX
DE HIV-type virus MWP5180/91 gag protein (cloned).
XX
KW Human immunodeficiency virus; HIV; antigen; detection; diagnosis;
retrovirus; vaccine; lymphocyte; reverse transcriptase.
OS
XX
PN HIV-type virus MWP-5180/91 (ECACC V92092318).
XX
PD EP591914-A.
XX
PD 13-APR-1994.
XX
PF 05-OCT-1993; 93EP-0116058.
XX
XX
XX 06-OCT-1992; 92DE-4233646.
PR 22-OCT-1992; 92DE-4235718.
PR 30-DEC-1992; 92DE-4244541.
PR 01-JUN-1993; 93DE-4318186.
XX
XX (BEHW ) BEHRINGWERKE AG.
PA
XX

```

PI Brunn VA, Eberle J, Gurtler IG, Hauser H-P, Knapp S;  
 XX WPI; 1994-120077/15.  
 DR N-PSDB; AAG58974.  
 XX  
 PT New HIV-type immune deficiency virus ECACC V 92092318 - and  
 PT deriv. cDNA or antigens, useful for diagnosing retroviral  
 PT infections and vaccines  
 XX  
 PS Disclosure: Fig 7; 73pp; German.  
 XX  
 CC WVP-5180/91 has been isolated from peripheral lymphocytes of a  
 CC patient from the Cameroons, with immune deficiency. It grows in the  
 CC same human cells as HIV-1; like HIV it produces an Mg-dependent  
 CC reverse transcriptase (RT), although this is 3 to 7 kD smaller in  
 CC than HIV-1, but more reactive than HIV-2, against p24-specific  
 CC antibodies and its gp41 transmembrane protein reacts with antibodies  
 CC in sera of African patients but not (or only weakly) with sera of  
 CC German patients.  
 CC The virus consists of RNA in a peg-shaped core made of p24 subunits  
 CC surrounded by an outer core of p17 and then a glycoprotein envelope  
 CC which, apart from host-cell derived lipids, comprises gp41 and  
 CC envelope protein gp120 (which can bind to the CD4 receptor).  
 CC Related viruses have at least 75% homology over the entire  
 CC genome with max. differences for the various regions LTR and NEF  
 CC 10%, POL 12%, GAG 14%, VIF 15% and ENV 22%.  
 CC  
 XX Sequence 498 AA;  
 SQ

Query Match 49.4%; Score 40; DB 15; Length 498;  
 Best Local Similarity 44.4%; Pred. No. 84;  
 Matches 8; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

OY 2 EERTPLTTAAXAPVYXNA 19  
 ||:| |:|:|  
 Db 119 eetsprqtsqnyplvtna 136

## RESULT 11

AAW93076  
 ID AAW93076 standard; Protein; 498 AA.  
 XX AAW93076;

19-MAY-1999 (first entry)

HIV isolate 5180 gag protein.

HIV-type retrovirus; WVP-5180/91; ECACC V 92092318; antigen assay kit;  
 detection: antibody; immune deficiency; vaccine.

Human immunodeficiency virus.

EP890642-A2.

13-JAN-1999.

05-OCT-1993; 93EP-0116058.

01-JUN-1993; 93DE-4318186.

06-OCT-1992; 92DE-4233646.

22-OCT-1992; 92DE-4235718.

30-DEC-1992; 92DE-4244541.

(DADE-) DADE BEHRING MARBURG GMBH.

Brunn AV, Eberle J, Gurtler IG, Hauser H, Knapp S;  
 WPI; 1999-072878/07.

New HIV-type retrovirus and corresponding cDNA, recombinant DNA and

PT antigen - used for detecting retro-viruses that cause immune  
 PT deficiency and to prepare vaccines  
 XX  
 XX

Example 11; Fig 7; 39pp; German.

This invention describes the isolation of a novel HIV-type retrovirus  
 CC called WVP-5180/91 (ECACC V 92092318). Antigens produced from this  
 CC product can be used in an assay kit for detecting antibodies against  
 CC viruses that cause immune deficiency, preferably where the assay is a  
 CC Western blot, ELISA or fluorescence immunoassay. WVP-5180/91, cDNA  
 CC and/or antigen can be used for detecting retroviruses that cause immune  
 CC deficiency and to prepare vaccines. This sequence represents an HIV  
 CC WVP 5180 gag protein.  
 CC  
 XX Sequence 498 AA;  
 SQ

Query Match 49.4%; Score 40; DB 20; Length 498;  
 Best Local Similarity 44.4%; Pred. No. 84;  
 Matches 8; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

OY 2 EERTPLTTAAXAPVYXNA 19  
 ||:| |:|:|  
 Db 119 eetsprqtsqnyplvtna 136

## RESULT 12

AAG39690  
 ID AAG39690 standard; Protein; 128 AA.  
 XX AAG39690;

18-OCT-2000 (first entry)

Arabidopsis thaliana protein fragment SEQ ID NO: 49147.

Protein identification: signal transduction pathway; metabolic pathway;  
 KW hybridisation assay; genetic mapping; gene expression control; promoter;  
 KW termination sequence.

Arabidopsis thaliana.

EP1033405-A2.

06-SEP-2000.

25-FEB-2000; 2000EP-0301439.

25-FEB-1999; 99US-0121825.

05-MAR-1999; 99US-0123180.

09-MAR-1999; 99US-0123548.

23-MAR-1999; 99US-0125788.

25-MAR-1999; 99US-0126264.

29-MAR-1999; 99US-0126785.

01-APR-1999; 99US-0127462.

06-APR-1999; 99US-0128234.

16-APR-1999; 99US-0128714.

19-APR-1999; 99US-0129845.

21-APR-1999; 99US-0130077.

23-APR-1999; 99US-0130449.

25-APR-1999; 99US-0130510.

28-APR-1999; 99US-0130891.

30-APR-1999; 99US-0131449.

04-MAY-1999; 99US-0132407.

06-MAY-1999; 99US-0132484.

06-MAY-1999; 99US-0132485.

06-MAY-1999; 99US-0132486.

07-MAY-1999; 99US-0132487.

11-MAY-1999; 99US-0132863.

14-MAY-1999; 99US-0134256.

14-MAY-1999; 99US-0134218.

14-MAY-1999; 99US-0134219.



PR 14-MAY-1999; 99US-0134221.  
 PR 14-MAY-1999; 99US-0134370.  
 PR 18-MAY-1999; 99US-0134768.  
 PR 19-MAY-1999; 99US-0134941.  
 PR 20-MAY-1999; 99US-0135124.  
 PR 21-MAY-1999; 99US-0135353.  
 PR 24-MAY-1999; 99US-0135629.  
 PR 25-MAY-1999; 99US-0136021.  
 PR 27-MAY-1999; 99US-0136392.  
 PR 28-MAY-1999; 99US-0136782.  
 PR 01-JUN-1999; 99US-0137222.  
 PR 03-JUN-1999; 99US-0137528.  
 PR 04-JUN-1999; 99US-0137502.  
 PR 07-JUN-1999; 99US-0137724.  
 PR 08-JUN-1999; 99US-0138094.  
 PR 10-JUN-1999; 99US-0138540.  
 PR 10-JUN-1999; 99US-0138847.  
 PR 14-JUN-1999; 99US-0139119.  
 PR 16-JUN-1999; 99US-0139452.  
 PR 16-JUN-1999; 99US-0139453.  
 PR 17-JUN-1999; 99US-0139492.  
 PR 18-JUN-1999; 99US-0139454.  
 PR 18-JUN-1999; 99US-0139455.  
 PR 18-JUN-1999; 99US-0139456.  
 PR 18-JUN-1999; 99US-0139457.  
 PR 18-JUN-1999; 99US-0139458.  
 PR 18-JUN-1999; 99US-0139459.  
 PR 18-JUN-1999; 99US-0139460.  
 PR 18-JUN-1999; 99US-0139461.  
 PR 18-JUN-1999; 99US-0139462.  
 PR 18-JUN-1999; 99US-0139463.  
 PR 18-JUN-1999; 99US-0139750.  
 PR 18-JUN-1999; 99US-0139763.  
 PR 21-JUN-1999; 99US-0139817.  
 PR 22-JUN-1999; 99US-0139899.  
 PR 23-JUN-1999; 99US-0140353.  
 PR 23-JUN-1999; 99US-0140354.  
 PR 24-JUN-1999; 99US-0140699.  
 PR 28-JUN-1999; 99US-0140829.  
 PR 29-JUN-1999; 99US-0140991.  
 PR 30-JUN-1999; 99US-0141287.  
 PR 01-JUL-1999; 99US-0141847.  
 PR 01-JUL-1999; 99US-0142154.  
 PR 02-JUL-1999; 99US-0142055.  
 PR 06-JUL-1999; 99US-0142390.  
 PR 08-JUL-1999; 99US-0142803.  
 PR 09-JUL-1999; 99US-0142920.  
 PR 12-JUL-1999; 99US-0142977.  
 PR 13-JUL-1999; 99US-0143542.  
 PR 14-JUL-1999; 99US-0143624.  
 PR 15-JUL-1999; 99US-0144005.  
 PR 16-JUL-1999; 99US-0144085.  
 PR 16-JUL-1999; 99US-0144086.  
 PR 19-JUL-1999; 99US-0144325.  
 PR 19-JUL-1999; 99US-0144331.  
 PR 19-JUL-1999; 99US-0144332.  
 PR 19-JUL-1999; 99US-0144333.  
 PR 19-JUL-1999; 99US-0144334.  
 PR 19-JUL-1999; 99US-0144335.  
 PR 20-JUL-1999; 99US-0144352.  
 PR 20-JUL-1999; 99US-0144632.  
 PR 20-JUL-1999; 99US-0144884.  
 PR 21-JUL-1999; 99US-0144814.  
 PR 21-JUL-1999; 99US-0145086.  
 PR 21-JUL-1999; 99US-0145088.  
 PR 22-JUL-1999; 99US-0145085.  
 PR 22-JUL-1999; 99US-0145087.  
 PR 22-JUL-1999; 99US-0145089.  
 PR 22-JUL-1999; 99US-0145192.  
 PR 23-JUL-1999; 99US-0145145.  
 PR 23-JUL-1999; 99US-0145224.  
 PR 23-JUL-1999; 99US-0145224.  
 PR 26-JUL-1999; 99US-0145276.

PR 27-JUL-1999; 99US-0145913.  
 PR 27-JUL-1999; 99US-0145918.  
 PR 27-JUL-1999; 99US-0145919.  
 PR 28-JUL-1999; 99US-0145951.  
 PR 02-AUG-1999; 99US-0146386.  
 PR 02-AUG-1999; 99US-0146388.  
 PR 03-AUG-1999; 99US-0146389.  
 PR 04-AUG-1999; 99US-0147038.  
 PR 04-AUG-1999; 99US-0147204.  
 PR 04-AUG-1999; 99US-0147302.  
 PR 05-AUG-1999; 99US-0147192.  
 PR 05-AUG-1999; 99US-0147260.  
 PR 06-AUG-1999; 99US-0147303.  
 PR 06-AUG-1999; 99US-0147416.  
 PR 09-AUG-1999; 99US-0147493.  
 PR 09-AUG-1999; 99US-0147935.  
 PR 10-AUG-1999; 99US-0148171.  
 PR 11-AUG-1999; 99US-0148319.  
 PR 13-AUG-1999; 99US-0148341.  
 PR 13-AUG-1999; 99US-0148565.  
 PR 16-AUG-1999; 99US-0149368.  
 PR 17-AUG-1999; 99US-0149175.  
 PR 18-AUG-1999; 99US-0149426.  
 PR 20-AUG-1999; 99US-0149722.  
 PR 20-AUG-1999; 99US-0149723.  
 PR 20-AUG-1999; 99US-0149929.  
 PR 23-AUG-1999; 99US-0149902.  
 PR 23-AUG-1999; 99US-0149930.  
 PR 25-AUG-1999; 99US-0150566.  
 PR 26-AUG-1999; 99US-0150884.  
 PR 27-AUG-1999; 99US-0151065.  
 PR 27-AUG-1999; 99US-0151086.  
 PR 30-AUG-1999; 99US-0151303.  
 PR 31-AUG-1999; 99US-0151438.  
 PR 01-SEP-1999; 99US-0151930.  
 PR 07-SEP-1999; 99US-0152363.  
 PR 10-SEP-1999; 99US-0153070.  
 PR 13-SEP-1999; 99US-0153758.  
 PR 15-SEP-1999; 99US-0154018.  
 PR 16-SEP-1999; 99US-0154039.  
 PR 20-SEP-1999; 99US-0154779.  
 PR 22-SEP-1999; 99US-0155139.  
 PR 23-SEP-1999; 99US-0155486.  
 PR 24-SEP-1999; 99US-0155659.  
 PR 28-SEP-1999; 99US-0156458.  
 PR 29-SEP-1999; 99US-0156596.  
 PR 04-OCT-1999; 99US-0157117.  
 PR 05-OCT-1999; 99US-0157753.  
 PR 06-OCT-1999; 99US-0157865.  
 PR 07-OCT-1999; 99US-0158029.  
 PR 08-OCT-1999; 99US-0158232.  
 PR 12-OCT-1999; 99US-0158369.  
 PR 13-OCT-1999; 99US-0159293.  
 PR 13-OCT-1999; 99US-0159294.  
 PR 13-OCT-1999; 99US-0159295.  
 PR 13-OCT-1999; 99US-0159329.  
 PR 14-OCT-1999; 99US-0159330.  
 PR 14-OCT-1999; 99US-0159331.  
 PR 14-OCT-1999; 99US-0159637.  
 PR 14-OCT-1999; 99US-0159658.  
 PR 18-OCT-1999; 99US-0159584.  
 PR 21-OCT-1999; 99US-0160741.  
 PR 21-OCT-1999; 99US-0160767.  
 PR 21-OCT-1999; 99US-0160768.  
 PR 21-OCT-1999; 99US-0160770.  
 PR 21-OCT-1999; 99US-0160814.  
 PR 21-OCT-1999; 99US-0160815.  
 PR 22-OCT-1999; 99US-0160980.  
 PR 22-OCT-1999; 99US-0160981.  
 PR 22-OCT-1999; 99US-0160989.  
 PR 25-OCT-1999; 99US-0161404.

PR 25-OCT-1999; 99US-0161405.  
 PR 25-OCT-1999; 99US-0161406.  
 PR 26-OCT-1999; 99US-0161359.  
 PR 26-OCT-1999; 99US-0161360.  
 PR 26-OCT-1999; 99US-0161361.  
 PR 28-OCT-1999; 99US-0161920.  
 PR 28-OCT-1999; 99US-0161992.  
 PR 28-OCT-1999; 99US-0161993.  
 PR 29-OCT-1999; 99US-0162142.

Query Match 48.1%; Score 39; DB 21; Length 128;  
 Best Local Similarity 66.7%; Pred. No. 28;  
 Matches 8; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 EKTPLTAXAP 14  
 ||| ||| |||  
 26 ekpssttasap 37

## RESULT 13

AAG91909  
 ID AAG91909 standard; Protein; 233 AA.

AC AAG91909;

DE 26-SEP-2001 (first entry)

C glutamicum protein fragment SEQ ID NO: 5663.

Coryneform bacterium; amino acid synthesis; vitamin; saccharide;  
 organic acid synthesis.

Corynebacterium glutamicum.

EP1108790-A2.

20-JUN-2001.

18-DEC-2000; 2000EP-0127688.

16-DEC-1999; 99JP-0377484.

07-APR-2000; 2000JP-0159162.

03-AUG-2000; 2000JP-0280988.

(KYOWA) KYOWA HAKKO KOGYO KK.

Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;  
 Tateishi N, Senoh A, Ikeda M, Ozaki A;

WPI; 2001-376931/40.

N-PSDB; AAH67128.

Novel polynucleotides derived from Coryneform bacteria, for identifying  
 mutation point of a gene, measuring expression of a gene, analysing  
 expression profile or pattern of a gene and identifying homologous gene

Claim 17; SEQ ID NO: 5663; 246pp + Sequence Listing; English.

The present invention provides a number of nucleotide and protein  
 sequences from the Coryneform bacterium Corynebacterium glutamicum. These  
 are useful for identifying the mutation point of a gene derived from a  
 mutant of coryneform bacterium, measuring expression amount and  
 analysing the expression profile or expression pattern of a gene derived  
 from coryneform bacterium, and identifying a homologue of a gene derived  
 from coryneform bacterium. Coryneform bacteria are useful for producing  
 amino acids, nucleic acids, vitamins, saccharides and organic acids,  
 particularly L-lysine. The present sequence is a protein described  
 in the exemplification of the invention.

Note: The sequence data for this patent did not form part of the printed  
 specification, but was obtained in electronic format directly from the  
 European Patent Office.

SO Sequence 233 AA;

Query Match 48.1%; Score 39; DB 22; Length 233;  
 Best Local Similarity 66.7%; Pred. No. 54;  
 Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 5 TPLTAXAPV 16  
 ||| ||| |||  
 35 tptstasapv 46

## RESULT 14

AAB94775  
 ID AAB94775 standard; Protein; 229 AA.

AC AAB94775;

DE 26-JUN-2001 (first entry)

Human protein sequence SEQ ID NO:15864.

Human; primer; detection; diagnosis; antisense therapy; gene therapy.

Homo sapiens.

EP1074617-A2.

07-FEB-2001.

28-JUL-2000; 2000EP-0116126.

29-JUL-1999; 99JP-0248036.

27-AUG-1999; 99JP-0300253.

11-JAN-2000; 2000JP-0118776.

02-MAY-2000; 2000JP-0183767.

09-JUN-2000; 2000JP-0241899.

(HELI-) HELIX RES INST.

Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
 Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

WPI; 2001-318749/34.

Primer sets for synthesizing polynucleotides, particularly the 5602  
 full-length cDNAs defined in the specification, and for the detection  
 and/or diagnosis of the abnormality of the proteins encoded by the  
 full-length cDNAs -

Claim 8; SEQ ID 15864; 2537pp + CD ROM; English.

The present invention describes primer sets for synthesizing 5602  
 full-length cDNAs defined in the specification. Where a primer set  
 comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
 to the complementary strand of a polynucleotide which comprises one of  
 the 5602 nucleotide sequences defined in the specification, where the  
 oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
 of an oligonucleotide comprising a sequence complementary to the  
 complementary strand of a polynucleotide which comprises a 5'-end  
 sequence and an oligonucleotide comprising a sequence complementary to a  
 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of  
 the 5'-end sequence/3'-end sequence is selected from those defined in  
 the specification. The primer sets can be used in antisense therapy and  
 in gene therapy. The primers are useful for synthesizing polynucleotides,  
 particularly full-length cDNAs. The primers are also useful for the  
 detection and/or diagnosis of the abnormality of the proteins encoded by  
 the full-length cDNAs. The primers allow obtaining of the full-length  
 cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
 AAH13633 to AAH1742 represent human cDNA sequences; AAB92446 to  
 AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632  
 represent oligonucleotides, all of which are used in the exemplification

CC of the present invention.  
XX  
SQ Sequence 229 AA;

	Query Match	46.9%	Score 38	DB 22	Length 229
	Best Local Similarity	39.3%	Pred. No. 78		
	Matches 11	Conservative	1	Mismatches 4	Indels 12
					Gaps 1
QY	3 EXPRLTTAA-----XAPVYXN	18			
	1111111111				
Db	109 expkltctclplspkprmdtclpvas	136			

RESULT 15  
AAG91455  
ID AAG91455 standard; Protein: 261 AA.

AC ANG91455;

26-SEP-2001 (first entry)

C glutamicum protein fragment SEQ ID NO: 5209

Coryneform bacterium; amino acid synthesis; vitamin; saccharide;  
organic acid synthesis.

OS *Corynebacterium glutamicum*.

PN EP1108790-A2.

PD 20-JUN-2001

18-DEC-2000; 2000EP-0127688.

PR 16-DEC-1999; 99JP-0377484.

03-AUG-2000; 2000JP-0280988

XX  
PA  
( KYOW ) KYOWA HAKKO KOGYO KI

XX Мзсавчт Н. Аг  
XX Мзсавчт Н. Аг

PI Tateishi N, Senoh A, Ikeda M, Ozaki A;

WPI; 2001-376931/40.

DR N FDD, 0000000000  
XX

PT Novel polynucleotide  
PT mutation point of

xx Novel polynucleotides derived from *Corynebacterium* bacteria, for identifying  
 pr mutation point of a gene, measuring expression of a gene, analysing  
 pr expression profile or pattern of a gene and identifying homologous gene  
 pr  
 Claim 17: SEQ ID NO: 5209; 246pp + Sequence Listing: English.

Callm 17: SEO ID NO: 5209; 246pp + Sequence Listing; English.

The present invention provides a number of nucleotide and protein sequences from the Coryneform bacterium *Corynebacterium glutamicum*. These sequences are useful for identifying the mutation point of a gene derived from a mutant of coryneform bacterium, measuring expression amount and analysing the expression profile or expression pattern of a gene derived from Coryneform bacterium, and identifying a homologue of a gene derived from coryneform bacterium. Coryneform bacteria are useful for producing amino acids, nucleic acids, vitamins, saccharides and organic acids, particularly L-lysine. The present sequence is a protein described in the exemplification of the invention.  
 Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from the European Patent Office.

Sequence 261 AA;

Query Match	46.98;	Score 38;	DB 22;	Length 261;
Best Local Similarity	70.08;	Pred. No. 90;		

	Matches	7;	Conservative	2;	Mismatches	1;	Indels	0;	Gaps	0;
Qy	6	PLTTAXAPV	15							
		:								
Db	149	plttasapv	158							

Search completed: February 28, 2002, 11:34:35  
Job time: 75 sec

Fri Mar 1 07:28:40 2002

us-09-359-426c-2.rag

Fri Mar 1 07:28:41 2002

us-09-359-426c-2.ra1

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 28, 2002, 11:33:53 : Search time 19.88 Seconds  
(without alignments)  
21.507 Million cell updates/sec

Title: US-09-359-426C-2  
Perfect score: 81  
Sequence: 1 XEEKPTLTAAKAPVXNA 19

Scoring table: BLOSUM62  
Gapop 10.0 , Capext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA: \*  
1: /cgn2\_6/pdata/2/1aa/5A.COMB.pep: \*  
2: /cgn2\_6/pdata/2/1aa/5B.COMB.pep: \*  
3: /cgn2\_6/pdata/2/1aa/6A.COMB.pep: \*  
4: /cgn2\_6/pdata/2/1aa/6B.COMB.pep: \*  
5: /cgn2\_6/pdata/2/1aa/PCRTUS.COMB.pep: \*  
6: /cgn2\_6/pdata/2/1aa/Backfilest.pep: \*

pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the total score being printed.  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	41	50.6	788	1	US-08-194-338-12 Sequence 12, Appl
2	40	49.4	498	1	US-08-470-202-59 Sequence 59, Appl
3	40	49.4	498	1	US-08-471-770-59 Sequence 59, Appl
4	40	49.4	498	2	US-08-468-059-59 Sequence 59, Appl
5	40	49.4	498	1	US-09-109-916-59 Sequence 20, Appl
6	37	45.7	463	2	US-08-679-635A-4 Sequence 8, Appl
7	37	45.7	463	2	US-08-679-635A-4 Sequence 8, Appl
8	36	44.4	58	3	US-08-943-173-8 Sequence 16, Appl
9	36	44.4	58	3	US-08-943-173-8 Sequence 16, Appl
10	36	44.4	71	3	US-08-943-173-2 Sequence 2, Appl
11	36	44.4	181	4	US-09-117-257-19 Sequence 19, Appl
12	36	44.4	181	4	US-09-117-257-19 Sequence 19, Appl
13	36	44.4	182	4	US-08-945-476-19 Sequence 48, Appl
14	36	44.4	355	2	US-08-458-555-2 Sequence 2, Appl
15	36	44.4	610	1	US-07-821-717B-6 Sequence 6, Appl
16	36	44.4	610	1	US-08-119-262B-6 Sequence 11, Appl
17	36	44.4	610	1	US-08-135-929A-11 Sequence 11, Appl
18	36	44.4	610	1	US-08-234-265A-11 Sequence 11, Appl
19	36	44.4	218	4	US-09-091-219-24 Sequence 142, App
20	35	43.2	267	4	US-08-818-112-142 Patent No. 5210183
21	35	43.2	344	6	US-09-032-315-4 Sequence 4, Appl
22	35	43.2	548	2	US-08-993-318A-4 Sequence 4, Appl
23	35	43.2	548	2	US-08-993-318A-4 Sequence 4, Appl
24	35	43.2	548	4	US-09-399-886-4 Sequence 4, Appl
25	35	43.2	548	4	US-09-396-260-4 Sequence 4, Appl
26	35	43.2	548	4	US-09-576-281-4 Sequence 4, Appl
27	35	43.2	662	1	US-07-841-651-4 Sequence 4, Appl

28	35	43.2	683	6	5210183-3	Patent No. 5210183
29	35	43.2	4551	3	US-09-320-878-1	Sequence 1, Appl
30	35	43.2	4613	4	US-09-105-537-31	Sequence 31, Appl
31	35	43.2	11877	4	US-09-105-537-6	Sequence 6, Appl
32	34	42.0	138	1	US-08-664-596B-22	Sequence 22, Appl
33	34	42.0	184	4	US-09-117-257-21	Sequence 21, Appl
34	34	42.0	184	4	US-08-945-476-21	Sequence 21, Appl
35	34	42.0	185	4	US-09-117-257-52	Sequence 52, Appl
36	34	42.0	263	1	US-07-927-071-2	Sequence 2, Appl
37	34	42.0	268	1	US-08-446-919A-2	Sequence 2, Appl
38	34	42.0	370	3	US-08-781-250-2	Sequence 2, Appl
39	34	42.0	447	1	US-08-450-360-2	Sequence 2, Appl
40	34	42.0	486	1	US-08-450-360-2	Sequence 2, Appl
41	34	42.0	542	3	US-08-968-563-16	Sequence 16, Appl
42	34	42.0	542	4	US-08-969-683A-16	Sequence 16, Appl
43	34	42.0	1063	1	US-08-093-453B-3	Sequence 3, Appl
44	34	42.0	1063	1	US-08-127-499A-8	Sequence 8, Appl
45	34	42.0	1063	1	US-08-482-847-8	Sequence 8, Appl

ALIGNMENTS

RESULT 1  
US-08-194-338-12 Application US/08194338  
Sequence 12, Application US/08194338  
Patent No. 5474898  
GENERAL INFORMATION:  
APPLICANT: Venter, John C.  
APPLICANT: Fraser, Claire M.  
APPLICANT: McCombie, William R.  
TITLE OF INVENTION: OCTOPAMINE RECEPTOR  
NUMBER OF SEQUENCES: 16  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Knobb, Martens, Olson and Bear  
Street: 620 Newport Center Drive, Sixteenth Floor  
City: Newport Beach  
STATE: CA  
COUNTRY: USA  
ZIP: 92660  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/194,338  
FILING DATE: 08-FEB-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/676,174  
FILING DATE: 28-MAR-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Israelsen, Ned A.  
REGISTRATION NUMBER: 29,655  
REFERENCE/DOCKET NUMBER: NIH01.001DVI  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 235-0176  
TELEFAX: (619) 235-0176  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 788 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: Internal  
US-08-194-338-12  
Query Match 50.6% Score 41: DB 1: Length 788:

Best Local Similarity 52.9%; Pred. No. 40;  
Matches 9; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 3 EEXPLTTAAAPVYXNA 19  
Db 602 EDOPTTAAAPLASAA 618

## RESULT 2

US-08-470-202-59  
Sequence 59, Application US/08470202  
Patent No. 5759808  
GENERAL INFORMATION:  
APPLICANT: Guerlier, Lutz G.  
APPLICANT: Ederle, Josef  
APPLICANT: Brunn, Albrecht V.  
APPLICANT: Knapp, Stefan  
APPLICANT: Hauser, Hans-Peter  
TITLE OF INVENTION: Retrovirus from the HIV Group and Its  
NUMBER OF SEQUENCES: 63  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &  
STREET: 1300 I Street, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20005  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/470,202  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/132,653  
FILING DATE: 05-OCT-1993  
APPLICATION NUMBER: DE P 42 33 646.5  
FILING DATE: 06-OCT-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: DE P 42 35 718.7  
FILING DATE: 22-OCT-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: DE P 42 44 541.8  
FILING DATE: 30-DEC-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: DE P 43 18 186.4  
FILING DATE: 01-JUN-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Michael J. Blake  
REGISTRATION NUMBER: 37,096  
REFERENCE/DOCKET NUMBER: 05495-0001-00000  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-408-4000  
TELEFAX: 202-408-4400  
INFORMATION FOR SEQ. ID NO: 59:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 498 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
FRAGMENT TYPE: internal  
US-08-470-202-59

Query Match 49.4%; Score 40; DB 1; Length 498;  
Best Local Similarity 44.4%; Pred. No. 35;  
Matches 8; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 2 EEXPLTTAAAPVYXNA 19  
Db 119 EETSPROTSONYPIVTNA 136

## RESULT 3

US-08-471-770-59  
Sequence 59, Application US/08471770  
Patent No. 5770427  
GENERAL INFORMATION:  
APPLICANT: Guerlier, Lutz G.  
APPLICANT: Ederle, Josef  
APPLICANT: Brunn, Albrecht V.  
APPLICANT: Knapp, Stefan  
APPLICANT: Hauser, Hans-Peter  
TITLE OF INVENTION: Retrovirus from the HIV Group and Its  
NUMBER OF SEQUENCES: 63  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &  
STREET: 1300 I Street, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20005  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/471,770  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/132,653  
FILING DATE: 05-OCT-1993  
APPLICATION NUMBER: DE P 42 33 646.5  
FILING DATE: 06-OCT-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: DE P 42 35 718.7  
FILING DATE: 22-OCT-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: DE P 42 44 541.8  
FILING DATE: 30-DEC-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: DE P 43 18 186.4  
FILING DATE: 01-JUN-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Carol P. Einaudi  
REGISTRATION NUMBER: 32,220  
REFERENCE/DOCKET NUMBER: 05495-0001-03000  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-408-4000  
TELEFAX: 202-408-4400  
INFORMATION FOR SEQ. ID NO: 59:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 498 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
FRAGMENT TYPE: internal  
US-08-471-770-59

Query Match 49.4%; Score 40; DB 1; Length 498;  
Best Local Similarity 44.4%; Pred. No. 35;  
Matches 8; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 2 EEXPLTTAAAPVYXNA 19



REFERENCE/DOCKET NUMBER: 2747-063-27  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 703-413-3000  
TELEFAX: 703-413-2220  
INFORMATION FOR SEQ ID NO: 20:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 456 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-624-125-20

Query Match 45.7%; Score 37; DB 1; Length 456;  
Best Local Similarity 72.7%; Pred. No. 1e+02;  
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 9 TTAAXAPVYXNA 19  
| | | | | | | |  
Db 336 TLAAPVYXNA 346

RESULT 7  
US-08-679-635A-4  
Sequence 4, Application US/08679635A  
Patent No. 5985643  
GENERAL INFORMATION:  
APPLICANT: Tomasz, Alexander  
APPLICANT: Delencastre, Herminia  
TITLE OF INVENTION: AUXILIARY GENES AND PROTEINS OF  
NUMBER OF SEQUENCES: 17  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: David A. Jackson, Esq.  
STREET: 411 Hackensack Ave, Continental Plaza, 4th  
FLOOR  
CITY: Hackensack  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 07601  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/679,635A  
FILING DATE: 10-JUL-1996  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Jackson Esq., David A.  
REGISTRATION NUMBER: 26,742  
REFERENCE/DOCKET NUMBER: 600-1-141  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201-487-5800  
TELEFAX: 201-343-1684  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 463 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
US-08-679-635A-4

Query Match 45.7%; Score 37; DB 2; Length 463;  
Best Local Similarity 66.7%; Pred. No. 1e+02;  
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
QY 8 TTAAXAPVYXNA 19

Db 382 TTAAPLVQTA 393  
| | | | | | | |

RESULT 8  
US-08-943-173-8  
Sequence 8, Application US/08943173  
Patent No. 6048538  
GENERAL INFORMATION:  
APPLICANT: Wang, Chang Yi  
APPLICANT: Shen, Fan  
APPLICANT: Chen, Pei De  
TITLE OF INVENTION: PEPTIDES DERIVED  
FROM THE NON-STRUCTURAL PROTEINS OF  
TITLE OF INVENTION: FOOT AND MOUTH DISEASE VIRUS AS  
DIAGNOSTIC REAGENTS  
NUMBER OF SEQUENCES: 16  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Maria C.H. Lin  
STREET: 345 Park Avenue  
CITY: New York  
STATE: NY  
COUNTRY: USA  
ZIP: 10154-0053  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WORDPERFECT 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/943,173  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Lin, Maria C.H.  
REGISTRATION NUMBER: 29,323  
REFERENCE/DOCKET NUMBER: 1151-4152  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-758-4800  
TELEFAX: 212-751-6849  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 19 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-943-173-8

Query Match 44.4%; Score 36; DB 3; Length 19;  
Best Local Similarity 53.3%; Pred. No. 4.1;  
Matches 8; Conservative 1; Mismatches 6; Indels 0; Gaps 0;  
QY 2 EKKPLTTAAXAPV 16  
| | | | | | | |  
Db 5 EKKPLTKAKAPV 19

RESULT 9  
US-08-943-173-16  
Sequence 16, Application US/08943173  
Patent No. 6048538  
GENERAL INFORMATION:  
APPLICANT: Wang, Chang Yi  
APPLICANT: Shen, Fan  
APPLICANT: Chen, Pei De  
TITLE OF INVENTION: PEPTIDES DERIVED  
FROM THE NON-STRUCTURAL PROTEINS OF  
TITLE OF INVENTION: FOOT AND MOUTH DISEASE VIRUS AS  
DIAGNOSTIC REAGENTS  
NUMBER OF SEQUENCES: 16  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Maria C.H. Lin





PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/589,711  
FILING DATE: 22-JAN-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/427,023  
FILING DATE: 24-APR-1995  
INFORMATION FOR SEQ ID NO: 19:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 181 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-945-476-19

Query Match 44.4%; Score 36; DB 4; Length 181;  
Best Local Similarity 47.1%; Pred. No. 53;  
Matches 8; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

OY 3 EKTPLTTAAAPVYXNA 19  
Db 139 EKTPTTAEGITITAKA 155

RESULT 13  
US-09-117-257-48  
Sequence 48, Application US/09117257  
Patent No. 6214355  
GENERAL INFORMATION:  
APPLICANT: Hook, Magnus  
APPLICANT: Guo, Betty  
APPLICANT: Hanson, Mark  
TITLE OF INVENTION: DbpA AND DbpB COMPOSITIONS AND METHODS OF USE  
FILE REFERENCE: 4210.000500  
CURRENT APPLICATION NUMBER: US/09/117,257  
CURRENT FILING DATE: 1998-07-22  
EARLIER APPLICATION NUMBER: PCT/US96/17081  
EARLIER FILING DATE: 1996-10-22  
EARLIER APPLICATION NUMBER: 08/589,711  
EARLIER FILING DATE: 1996-01-22  
EARLIER APPLICATION NUMBER: 08/427,023  
NUMBER OF SEQ ID NOS: 66  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 48  
LENGTH: 182  
TYPE: PRT  
ORGANISM: Borrelia afzelii  
US-09-117-257-48

Query Match 44.4%; Score 36; DB 4; Length 182;  
Best Local Similarity 47.1%; Pred. No. 53;  
Matches 8; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

OY 3 EKTPLTTAAAPVYXNA 19  
Db 139 EKTPTTAEGITITAKA 155

RESULT 14  
US-08-458-555-2  
Sequence 2, Application US/08458555  
Patent No. 5840689  
GENERAL INFORMATION:  
APPLICANT: Daniloff, Joanne K  
APPLICANT: McIntjes, Elmarie  
TITLE OF INVENTION: Method for Stimulating the Regrowth of  
NUMBER OF INVENTION: Neurons  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: John H. Runnels  
STREET: P. O. Box 2471

CITY: Baton Rouge  
STATE: LA  
COUNTRY: USA  
ZIP: 70821-2471  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/458,555  
FILING DATE:  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/146,734  
FILING DATE: 29-OCT-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Runnels, John H  
REGISTRATION NUMBER: 33451  
REFERENCE/DOCKET NUMBER: Attorney No. 5840689 9312  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 504 387-3221  
TELEFAX: 504 346-8049  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 355 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
FRAGMENT TYPE: N-terminal  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
US-08-458-555-2

Query Match 44.4%; Score 36; DB 2; Length 355;  
Best Local Similarity 61.5%; Pred. No. 1,1e+02;  
Matches 8; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 3 EKTPLTTAAAPV 15  
Db 308 ETPPLEPEKAPV 320

RESULT 15  
US-07-821-717B-6  
Sequence 6, Application US/07821717B  
Patent No. 5298239  
GENERAL INFORMATION:  
APPLICANT: Miller, Jonathan L.  
APPLICANT: Cunningham, David  
APPLICANT: Lyle, Vicki A.  
APPLICANT: Finch, Clara N.  
TITLE OF INVENTION: MUTATIONS RENDERING PLATELET  
TITLE OF INVENTION: GLYCOPROTEIN Ib ALPHA LESS REACTIVE  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Nixon, Hargrave, Devans & Doyle  
STREET: Clinton Square, P.O. Box 1051  
CITY: Rochester  
STATE: New York  
COUNTRY: USA  
ZIP: 14603  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/821,717B  
FILING DATE: 15-JAN-1992

CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Timain, Susan J.  
REGISTRATION NUMBER: 34,103  
REFERENCE/DOCKET NUMBER: 20884/21  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (716) 263-1636  
TELEFAX: (716) 263-1600  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 610 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PUBLICATION INFORMATION:  
AUTHORS: Lopez, Jose A.  
AUTHORS: Chung, Dominic W.  
AUTHORS: Fujikawa, Kazuo  
AUTHORS: Hagen, Frederick S.  
AUTHORS: Papayannopoulou, Thalia  
AUTHORS: Roth, Gerald J.  
TITLE: Cloning of the alpha chain of human  
TITLE: platelet glycoprotein Ib: A transmembrane protein with homology  
TITLE: to leucine-rich alpha-2-glycoprotein  
JOURNAL: Proc. Natl. Acad. Sci. U.S.A.  
VOLUME: 84  
PAGES: 5615-5619  
DATE: AUG-1987  
RELEVANT RESIDUES IN SEQ ID NO: 6: FROM 0 TO 610  
PUBLICATION INFORMATION:  
AUTHORS: Zimmerman, Theodore S.  
AUTHORS: Ruggeri, Zaverio M.  
AUTHORS: Houghten, Richard A.  
AUTHORS: Vincete, Vincete  
AUTHORS: Mohl, Hiroshi  
TITLE: proteolytic fragments and synthetic  
TITLE: peptides that block the binding of von Willebrand factor to the  
TITLE: platelet membrane glycoprotein Ib  
DOCUMENT NUMBER: EP 0 317 278 A2  
FILING DATE: 16-NOV-1988  
PUBLICATION DATE: 24-MAY-1989  
RELEVANT RESIDUES IN SEQ ID NO: 6: FROM 0 TO 293  
US-07-821-717B-6

Query Match 44.4% Score 36; DB 1; Length 610;  
Best Local Similarity 42.9% Pred. No. 2.1e+02;  
Matches 6; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

3 EKTPLTTAAXAPV 16  
409 EPTPIPTIATSPTI 422

Search completed: February 28, 2002, 11:33:53  
Job time: 29 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: February 28, 2002, 11:35:08 ; Search time 23.26 Seconds  
(without alignments)  
62.223 Million cell updates/sec

Title: US-09-359-426C-2

Perfect score: 81

Sequence: 1 XEKTPLTTAAXAPVYXNA 19

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

1: PIR:68:\*  
2: PIR:1:\*  
3: PIR:2:\*  
4: PIR:3:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	63	77.8	482	2	B83113 catalase PA4236 [1
2	53	65.4	484	2	A58663 catalase (EC 1.11.
3	46	56.8	262	2	T33408 hypothetical prote
4	45	55.6	128	2	JC1273 ribosomal protein
5	44	54.3	427	2	F85436 hypothetical prote
6	43	53.1	483	2	S37055 catalase (EC 1.11.
7	42	51.9	105	2	A44639 catalase (EC 1.11.
8	42	51.9	321	2	T08462 hypothetical prote
9	42	51.9	436	2	T46107 copper chaperone h
10	41	50.6	132	2	T50779 catalase (EC 1.11.
11	41	50.6	482	2	S60757 catalase (EC 1.11.
12	40	49.4	171	2	T31478 hypothetical prote
13	40	49.4	252	2	H72469 muscarinic acetyl
14	40	49.4	788	2	S05661 hypothetical prote
15	40	49.4	1360	2	T31674 hypothetical prote
16	39.5	48.8	405	2	J02147 OHPI protein - mal
17	39	48.1	486	2	S10772 2-hydroxyxymuconic s
18	39	48.1	527	2	S46088 hypothetical prote
19	39	48.1	1353	2	T26301 hypothetical prote
20	38.5	47.5	1658	2	D75489 growth-blocking pe
21	38	46.9	143	2	S68226 cytochrome c1, hem
22	38	46.9	307	2	T40089 phospholipase D (E
23	38	46.9	307	2	A51125 hypothetical prote
24	38	46.9	348	2	S40750 hypothetical prote
25	38	46.9	532	2	S40983 peptidase (M3 fami
26	38	46.9	573	2	F81313 hypothetical prote
27	38	46.9	684	2	T25603 probable translati
28	38	46.9	695	2	D71283 hypothetical prote
29	38	46.9	1679	2	S48385

30	37.5	46.3	200	2	G86194 hypothetical prote
31	37	45.7	71	2	F72332 hypothetical prote
32	37	45.7	116	2	C83492 hypothetical prote
33	37	45.7	328	2	T01225 hypothetical prote
34	37	45.7	347	2	T35518 probable secreted
35	37	45.7	463	2	S72992 probable phosphory
36	37	45.7	488	2	T42038 catalase (EC 1.11.
37	37	45.7	500	2	S72505 lycopene beta-cycl
38	37	45.7	500	2	S72506 lycopene beta-cycl
39	37	45.7	500	2	S66349 lycopene beta-cycl
40	37	45.7	500	2	S66350 aldehyde dehydroge
41	37	45.7	511	1	S31308 hypothetical prote
42	37	45.7	526	2	T23779 hypothetical prote
43	37	45.7	534	2	S62572 probable ABC-type
44	37	45.7	539	2	T36524 mucin FIW-C.1 - Af
45	37	45.7	662	2	A45155

#### ALIGNMENTS

##### RESULT 1

B83113 catalase PA4236 [Imported] - Pseudomonas aeruginosa (strain PA01)

C:Species: Pseudomonas aeruginosa

C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000

C:Accession: B83113

R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.;

adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L

.. Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000

A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa

A:Reference number: A82950; MUID:20437337

A:Accession: B83113

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-482 <STD>

A:Cross-References: GB:AE004841; GB:AE004091; NID:9950451; PIDN:AC07624.1; GSPDB:GN

A:Experimental source: strain PA01

C:Genetics:

A:Gene: katA; PA4236

C:Superfamily: catalase

Query Match 77.8%; Score 63; DB 2; Length 482;

Best local similarity 82.4%; Pred. No. 0.0032;

Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 EKTPLTTAAXAPVYXN 18

DB 2 EKTPLTTAAXAPVYXN 18

##### RESULT 2

A58663 catalase (EC 1.11.1.6) [validated] - Proteus mirabilis

C:Species: Proteus mirabilis

A:Variety: strain P1, peroxide resistant

C:Date: 19-Nov-1997 #sequence\_revision 21-Nov-1997 #text\_change 15-Sep-2000

C:Accession: A58663; B58663

R:Buzy, A.; Brachi, V.; Sterliades, R.; Chroboczek, J.; Thibault, P.; Gagnon, J.; Jo

U. Protein Chem. 14, 59-72, 1995

A:Title: Complete amino acid sequence of Proteus mirabilis PR catalase. Occurrence of

A:Reference number: A58663; MUID:95305957

A:Accession: A58663

A:Molecule type: protein

A:Residues: 1-484 <B021>

A:Accession: B58663

A:Molecule type: DNA

A:Residues: 1-305, 'AE' <B022>

R:Gouet, P.; Jouve, H.M.; Dideberg, O.

submitted to the Brookhaven Protein Data Bank, June 1996

A:Reference number: A67899; PDB:2CAE

A:Contents: annotation; X-ray crystallography, 2.2 angstroms, residues 3-475  
 R:Gouet, P.; Jouve, H.M.; Hajdu, J.  
 Submitted to the Brookhaven Protein Data Bank, June 1996  
 A:Reference number: A67900; PDB:2CAG  
 A:Contents: annotation; X-ray crystallography, 2.7 angstroms, residues 3-475  
 A:Note: compound I  
 R:Gouet, P.; Jouve, H.M.; Hajdu, J.  
 Submitted to the Brookhaven Protein Data Bank, June 1996  
 A:Reference number: A67901; PDB:2CAG  
 A:Contents: annotation; X-ray crystallography, 2.7 angstroms, residues 3-475  
 A:Note: compound II, dithiothreitol reduced compound I  
 R:Gouet, P.; Jouve, H.M.; Dideberg, O.  
 Submitted to the Brookhaven Protein Data Bank, July 1996  
 A:Reference number: A67902; PDB:2CAH  
 A:Contents: annotation; X-ray crystallography, 2.7 angstroms, residues 3-475  
 A:Note: native Fe(III) with NADPH  
 R:Gouet, P.; Jouve, H.M.; Dideberg, O.  
 Mol. Biol. 249, 933-954, 1995  
 A:Title: Crystal structure of Proteus mirabilis PR catalase with and without bound NADPH  
 A:Reference number: A58664; M0ID:95311317  
 A:Contents: annotation; X-ray crystallography, 2.2 angstroms  
 C:Complex: homotetramer  
 C:Function:  
 A:Description: catalyzes the conversion of two of molecules of hydrogen peroxide to two  
 A:Note: this enzyme has a tightly bound NADPH cofactor  
 C:Superfamily: catalase  
 C:Keywords: Chromoprotein; heme; homotetramer; iron; metalloprotein; NADP; oxidoreductase  
 F:53/Modified site: methionine sulfone (Met) #status experimental  
 F:54,93,127/Active site: His, Ser, Asn #status predicted  
 F:337/Binding site: heme iron (Tyr) (axial ligand) #status experimental

Query Match 65.4%; Score 53; DB 2; Length 484;  
 Best Local Similarity 75.0%; Pred. No. 0.17;  
 Matches 12; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 3 EKTPLTTAAXAPVYN 18  
 || ||||| |||||  
 DB 2 EKKKLTAAAGAPVDN 17

RESULT 3  
 T33408  
 Hypothetical protein H34124.2 - Caenorhabditis elegans (fragment)  
 A:Species: Caenorhabditis elegans  
 A:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999  
 C:Accession: T33408  
 R:Latreille, P.; Wamsley, P.; O'Brien, D.  
 submitted to the EMBL Data Library, July 1998  
 A:Description: The sequence of C. elegans cosmid H34124.  
 A:Reference number: Z21340  
 A:Accession: T33408  
 A:Status: preliminary; translated from GB/EMBL/DBD  
 A:Molecule type: DNA  
 A:Residues: 1-262 <LAT>  
 A:Cross-references: EMBL:AF078784; PIDN:AAC26925.1; GSPDB:GN00021; CESP:H34124.2  
 C:Experimental source: strain Bristol N2; clone H34124  
 C:Genetics:  
 A:Gene: CESP:H34124.2  
 A:Map position: 3  
 A:Note: intron positions not resolved (incomplete sequence)

Query Match 56.8%; Score 46; DB 2; Length 262;  
 Best Local Similarity 56.2%; Pred. No. 1.4;  
 Matches 9; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

OY 4 KTLPLTAAAPVYN 19  
 ||||| : || : ||  
 DB 190 KTLPLTSSAARVINA 205

RESULT 4

JC1273  
 Ribosomal protein L7/L12 - Streptomyces antibioticus  
 C:Species: Streptomyces antibioticus  
 C:Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 13-Aug-1999  
 C:Accession: JC1273  
 R:Parra, F.; Blanco, G.; Alonso, J.M.M.; Balbin, M.; Mendez, C.; Salas, J.A.  
 Gene 118, 127-129, 1992  
 A:Title: Cloning and sequence of a gene encoding the L7/L12 ribosomal protein equivalent  
 A:Reference number: JC1273; M0ID:92380478  
 A:Accession: JC1273  
 A:Molecule type: DNA  
 A:Residues: 1-128 <PAR>  
 A:Cross-references: GB:M89911; NID:g153436; PIDN:AAA26811.1; PID:g153438  
 C:Superfamily: Escherichia coli ribosomal protein L12  
 C:Keywords: protein biosynthesis; ribosome

Query Match 55.6%; Score 45; DB 2; Length 128;  
 Best Local Similarity 66.7%; Pred. No. 0.98;  
 Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 2 EKTPLTTAAXAPV 16  
 || : || ||||  
 DB 28 EKEFDVTAATAAPV 42

RESULT 5  
 F85436  
 Hypothetical protein AT4g36970 [imported] - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 16-Feb-2001  
 C:Accession: F85436  
 R:Anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Sp  
 Nature 402, 769-777, 1999  
 A:Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.  
 A:Reference number: A85001; M0ID:20083488  
 A:Accession: F85436  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-427 <STO>  
 A:Cross-references: GB:NC\_001268; NID:g7270646; PIDN:CAB80363.1; GSPDB:GN00140  
 C:Genetics:  
 A:Gene: AT4g36970  
 A:Map position: 4

Query Match 54.3%; Score 44; DB 2; Length 427;  
 Best Local Similarity 60.0%; Pred. No. 5.5;  
 Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 2 EKTPLTTAAXAPV 16  
 || : |||| ||||  
 DB 208 EKAEMTTAMQSPV 222

RESULT 6  
 S37055  
 catalase (EC 1.11.1.6) alpha-2 chain - Streptomyces violaceus  
 C:Species: Streptomyces violaceus  
 C:Date: 10-Dec-1993 #sequence\_revision 26-May-1995 #text\_change 04-Mar-2000  
 C:Accession: S37055  
 R:Facey, S.; van Pee, K.H.; Vining, L.C.  
 submitted to the EMBL Data Library, August 1993  
 A:Reference number: S37055  
 A:Accession: S37055  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-483 <FAC>  
 A:Cross-references: EMBL:X74791; NID:g9397888; PIDN:CAA52796.1; PID:g581780  
 A:Note: the source is designated as Streptomyces venezuelae  
 C:Genetics:  
 A:Start codon: GTG  
 C:Superfamily: catalase

C:Keywords: chromoprotein; heme; iron; metalloprotein; oxidoreductase  
 F:54,93,127/Active site: His, Ser, Asn #status predicted  
 F:337/Binding site: heme iron (Tyr) (axial ligand) #status predicted

Query Match 51.9% Score 42; DB 2; Length 483;  
 Best Local Similarity 69.2%; Pred. No. 9.4;  
 Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 OY 6 PLTTAAAPVYXN 18  
 DB 5 PLTTAGAPVADN 17

RESULT 7  
 A44639  
 catalase (EC 1.11.1.6) - Streptomyces coelicolor (fragments)  
 C:Species: Streptomyces coelicolor  
 C:Date: 07-Oct-1994 #sequence\_revision 07-Oct-1994 #text\_change 31-Oct-1997  
 C:Accession: A44639  
 R:Walker, G.  
 submitted to the Protein Sequence Database, September 1994  
 :Reference number: A44639  
 :Accession: A44639  
 :Status: preliminary  
 A:Molecule type: protein  
 A:Residues: 1-105 <NAL>  
 C:Superfamily: catalase  
 C:Keywords: oxidoreductase

Query Match 51.9% Score 42; DB 2; Length 105;  
 Best Local Similarity 69.2%; Pred. No. 2.6;  
 Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 OY 6 PLTTAAAPVYXN 18  
 DB 7 PLTTAGAPVADN 19

RESULT 8  
 T08462  
 hypothetical protein F2206.250 - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 22-Oct-1999  
 C:Accession: T08462  
 R:Queller, F.; Purnelle, B.; Boutry, M.; Goffeau, A.; Salanoubat, M.; Mewes, H.W.; Mayer  
 submitted to the Protein Sequence Database, May 1999  
 A:Reference number: 216420  
 A:Accession: T08462  
 A:Molecule type: DNA  
 A:Residues: 1-321 <QUR>  
 A:Cross-references: EMBL:AL050300; GSPDB:GN00061; ATSP:F2206.250  
 A:Experimental source: cultivar Columbia; BAC clone F2206  
 C:Genetics:  
 A:Gene: ATSP:F2206.250  
 A:Map position: 3

Query Match 51.9% Score 42; DB 2; Length 321;  
 Best Local Similarity 50.0%; Pred. No. 9;  
 Matches 9; Conservative 1; Mismatches 8; Indels 0; Gaps 0;  
 OY 2 EEKTPPLTTAAAPVYXNA 19  
 DB 291 EEKTPVEKKTGCVPVYKKA 308

RESULT 9  
 T46107  
 hypothetical protein T25B15.140 - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 04-Feb-2000 #sequence\_revision 04-Feb-2000 #text\_change 04-Feb-2000

C:Accession: T46107  
 R:Alcaraz, J.P.; Clabault, G.; Collet, A.; Maché, R.; Mewes, H.W.; Lemcke, K.; Mayer,  
 submitted to the Protein Sequence Database, January 2000  
 A:Reference number: 223021  
 A:Accession: T46107  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-436 <ALC>  
 A:Cross-references: EMBL:AL132972  
 A:Experimental source: cultivar Columbia; BAC clone T25B15  
 C:Genetics:  
 A:Map position: 3  
 A:Introns: 418/1  
 A:Note: T25B15.140

Query Match 51.9% Score 42; DB 2; Length 436;  
 Best Local Similarity 50.0%; Pred. No. 13;  
 Matches 9; Conservative 1; Mismatches 8; Indels 0; Gaps 0;  
 OY 2 EEKTPPLTTAAAPVYXNA 19  
 DB 395 EEKTPVEKKTGCVPVYKKA 412

RESULT 10  
 T50779  
 copper chaperone homolog CCH [Imported] - rice  
 C:Species: Oryza sativa (rice)  
 C:Date: 21-Jul-2000 #sequence\_revision 21-Jul-2000 #text\_change 21-Jul-2000  
 C:Accession: T50779  
 R:Himelblau, E.; Mira, H.; Lin, S.-J.; Culotta, V.C.; Penarrubia, L.; Amasino, R.M.  
 Plant Physiol. 117, 1227-1234, 1998  
 A:Title: Identification of a functional homolog of the yeast copper homeostasis gene  
 A:Reference number: 224450; MUID:9701579  
 A:Accession: T50779  
 A:Status: preliminary; translated from GR/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-132 <HIM>  
 A:Cross-references: EMBL:AF198626; PIDN:AAF15285.1

Query Match 50.6% Score 41; DB 2; Length 132;  
 Best Local Similarity 50.0%; Pred. No. 5;  
 Matches 9; Conservative 1; Mismatches 8; Indels 0; Gaps 0;  
 OY 2 EEKTPPLTTAAAPVYXNA 19  
 DB 97 EAAPPTTAENPAVIAA 114

RESULT 11  
 S60757  
 catalase (EC 1.11.1.6) - Bordetella pertussis  
 C:Species: Bordetella pertussis  
 C:Date: 27-Apr-1996 #sequence\_revision 13-Mar-1997 #text\_change 04-Mar-2000  
 C:Accession: S60757  
 R:Deshaizer, D.; Wood, G.E.; Friedman, R.L.  
 Mol. Microbiol. 14, 123-130, 1994  
 A:Title: Molecular characterization of catalase from Bordetella pertussis: Identifica  
 A:Reference number: S60757; MUID:95131725  
 A:Accession: S60757  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-482 <DBS>  
 A:Cross-references: EMBL:U07800; MID:g494943; PIDN:AAA18481.1; PID:g494944  
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, March 1994  
 C:Genetics:  
 A:Gene: cata  
 C:Superfamily: catalase  
 C:Keywords: chromoprotein; heme; iron; metalloprotein; oxidoreductase  
 F:57,96,130/Active site: His, Ser, Asn #status predicted  
 F:340/Binding site: heme iron (Tyr) (axial ligand) #status predicted

Query Match 50.6%; Score 41; DB 2; Length 482;  
 Best Local Similarity 75.0%; Pred. No. 21;  
 Matches 9; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 7 LTTAAAPVYXN 18  
 |||||  
 Db 9 LTTAGAPVADN 20

RESULT 12  
 T31478  
 hypothetical protein F56F12.1 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C>Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999  
 Accession: T31478  
 Muscouph, R.  
 submitted to the EMBL Data Library, October 1999  
 A:Reference number: Z21039  
 A:Accession: T31478  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-171 <WIL>  
 A:Cross-references: EMBL:Z82273; PIDN:CAB54978.1; GSPDB:GN00022; CESP:F56F12.1  
 A:Experimental source: clone F56F12  
 C:Genetics:  
 A:Gene: CESP:F56F12.1  
 A:Map position: 4  
 A:Introns: 16/2; 50/1

Query Match 49.4%; Score 40; DB 2; Length 171;  
 Best Local Similarity 53.3%; Pred. No. 10;  
 Matches 8; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

OY 5 TPTTAAAPVYXNA 19  
 |||||  
 Db 73 TPTTAAAPVYXNA 87

RESULT 13  
 H72469  
 hypothetical protein APE2401 - Aeropyrum pernix (strain K1)  
 C:Species: Aeropyrum pernix  
 C>Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 20-Aug-1999  
 Accession: H72469  
 R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takei,  
 awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; K  
 DNA Res. 6, 83-101, 1999  
 A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyr  
 A:Reference number: A72450; MUID:99310339  
 A:Accession: H72469  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-252 <KAM>  
 A:Cross-references: DDBJ:AP000064; NID:95105945; PIDN:BA81416.1; PID:dl045202; PID:9510  
 A:Experimental source: strain K1  
 C:Genetics:  
 A:Gene: APE2401

Query Match 49.4%; Score 40; DB 2; Length 252;  
 Best Local Similarity 88.9%; Pred. No. 15;  
 Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 6 PLTTAAAPVYXN 14  
 |||||  
 Db 64 PLTTAAAPVYXN 72

RESULT 14  
 S05661

muscarinic acetylcholine receptor - fruit fly (Drosophila melanogaster)  
 C:Species: Drosophila melanogaster  
 C>Date: 30-Jun-1991 #sequence\_revision 30-Jun-1991 #text\_change 15-Oct-1999  
 C:Accession: S05661; A36191  
 R:Onai, T.; Fitzgerald, M.G.; Atakawa, S.; Gocayne, J.D.; Urquhart, D.A.; Hall, L.M.;  
 FEBS Lett. 255, 219-225, 1989  
 A:Title: Cloning, sequence analysis and chromosome localization of a Drosophila musca  
 A:Reference number: S05661; MUID:90005981  
 A:Accession: S05661

A:Molecule type: mRNA  
 A:Residues: 1-788 <ONA>  
 A:Cross-references: EMBL:M23412; NID:9157829; PID:9157830  
 R:Shapiro, R.A.; Wakimoto, B.T.; Subers, E.M.; Nathanson, N.M.  
 Proc. Natl. Acad. Sci. U.S.A. 86, 9039-9043, 1989  
 A:Title: Characterization and functional expression in mammalian cells of genomic and  
 A:Reference number: A36191; MUID:90046926  
 A:Accession: A36191  
 A:Molecule type: mRNA  
 A:Residues: 81-154, 'T', 156-198, 'N', 200-215, 'G', 217-227, 229-330, 'P', 332-426, 'YATPTTET  
 A:Cross-references: GB:M27495  
 A:Experimental source: strain Oregon R  
 C:Genetics:  
 A:Gene: FLYBase:macr-60C  
 A:Map position: 2R 60C5-C8  
 A:Cross-references: FLYBase:FBgn0000037  
 C:Keywords: G protein-coupled receptor; glycoprotein; neurotransmitter receptor; tran  
 F:104-127/Domain: transmembrane #status predicted <TM1>  
 F:141-161/Domain: transmembrane #status predicted <TM2>  
 F:179-200/Domain: transmembrane #status predicted <TM3>  
 F:220-243/Domain: transmembrane #status predicted <TM4>  
 F:268-289/Domain: transmembrane #status predicted <TM5>  
 F:702-722/Domain: transmembrane #status predicted <TM6>  
 F:739-758/Domain: transmembrane #status predicted <TM7>

Query Match 49.4%; Score 40; DB 2; Length 788;  
 Best Local Similarity 52.9%; Pred. No. 54;  
 Matches 9; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

OY 3 EKTPLTTAAAPVYXNA 19  
 |||||  
 Db 602 EDGPTTTAAAPVYXNA 618

RESULT 15  
 T31674  
 hypothetical protein 1 - fruit fly (Drosophila simulans) retrotransposon ninja (fragm  
 C:Species: Drosophila simulans  
 C>Date: 02-Sep-2000 #sequence\_revision 02-Sep-2000 #text\_change 17-Nov-2000  
 C:Accession: T31674  
 R:Ogura, K.; Takechi, S.; Nakayama, T.; Yamamoto, M.  
 Genes Genet. Syst. 71, 1-8, 1996  
 A:Title: Molecular structure of the transposable element ninja in Drosophila simulans  
 A:Reference number: Z21057; MUID:96252483  
 A:Accession: T31674  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-1360 <OGU>  
 A:Cross-references: EMBL:D83207; NID:dl068446; PID:dl020552; PIDN:BA19771.1  
 C:Genetics:  
 A:Cross-references: FLYBase:FBgn0015168  
 A:Mobile element: retrotransposon ninja

Query Match 49.4%; Score 40; DB 2; Length 1360;  
 Best Local Similarity 47.1%; Pred. No. 98;  
 Matches 8; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

OY 3 EKTPLTTAAAPVYXNA 19  
 |||||  
 Db 101 EATPTTASQOPASSGA 117



Fri Mar 1 07:28:45 2002

Search completed: February 28, 2002, 11:35:09  
Job time: 105 sec

us-09-359-426c-2.rpt

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 28, 2002, 11:39:52 ; search time 12.98 seconds  
(without alignments)  
53.670 Million cell updates/sec

Title: US-09-359-426c-2

Perfect score: 81  
Sequence: 1 XEKPPLPTAAKAPVYXNA 19

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database: SwissProt\_39:\*

pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	63	77.8	482	1	CATP_PSEAE
2	53	65.4	484	1	CATP_PSEAE
3	45	55.6	128	1	RL7_STRAT
4	43	53.1	483	1	BCA_STRAT
5	41	50.6	482	1	CATP_PSEAE
6	40	49.4	722	1	ACM1_DROME
7	39	48.1	143	1	H2B_AGABI
8	39	48.1	486	1	DMPC_PSEAE
9	39	48.1	527	1	YB64_YEAST
10	39	48.1	560	1	VNIC_INBLE
11	39	48.1	815	1	GRB_MYXMA
12	38	46.9	143	1	GBP_PSEAE
13	38	46.9	307	1	PLD_CORPS
14	38	46.9	479	1	CATP_PSEAE
15	38	46.9	695	1	EGF1_TREPA
16	38	46.9	747	1	YMAH_CAEEL
17	38	46.9	1679	1	YIO9_YEAST
18	37	45.7	366	1	FTZ1_PYRMO
19	37	45.7	442	1	SDC3_MOUSE
20	37	45.7	482	1	CATP_VIBFI
21	37	45.7	498	1	LCYB_CAPAN
22	37	45.7	500	1	LCYB_LYCES
23	37	45.7	500	1	LCYB_TOBAC
24	37	45.7	500	1	PD32_ANOST
25	37	45.7	511	1	DHAY_YEAST
26	37	45.7	534	1	YAUJ_SCHPO
27	37	45.7	662	1	MUC1_XENLA
28	37	45.7	946	1	FTSK_CAMJE
29	37	45.7	1034	1	GCSP_FILAN
30	37	45.7	1064	1	ISKS_HUMAN
31	37	45.7	1567	1	ICEN_XANCT
32	37	45.7	3390	1	POLG_DEN3
33	36.5	45.1	235	1	CAV1_CAEEL

34	36	44.4	158	1	HUNB_DROME
35	36	44.4	214	1	DAMX_SPRMA
36	36	44.4	332	1	ANT_BPPI
37	36	44.4	501	1	LCYB_ARATH
38	36	44.4	553	1	SPAK_RAT
39	36	44.4	567	1	ODP2_HAEIN
40	36	44.4	626	1	GPBA_HUMAN
41	36	44.4	1128	1	BEM3_YEAST
42	36	44.4	2332	1	POLG_PMDVA
43	36	44.4	205	1	POLG_PMDVO
44	35	43.2	205	1	YJ11_MYCTU
45	35	43.2	222	1	H15_MOUSE

## ALIGNMENTS

RESULT 1  
CATP\_PSEAE STANDARD: PRT: 482 AA.  
ID CATP\_PSEAE  
AC 052762, 2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE CATALASE (EC 1.11.1.6).  
GN KATA OR PA4236.  
OS Pseudomonas aeruginosa.  
OC Bacteria: Proteobacteria: gamma subdivision: Pseudomonadaceae;  
OC Pseudomonas.  
OC NCBI\_TaxID=287;  
OX (1)  
RN SEQUENCE FROM N.A.  
RP STRAIN=PRDI;  
RC MEDLINE=99296583; PubMed=10368148;  
RX Ma J.-F., Ochser U.A., Klotz M.G., Nanayakkara V.K., Howell M.L.,  
RA Johnson Z., Posey J.E., Vasil M.L., Monaco J.J., Hassett D.J.,  
RT "Bacterioferitin A modulates catalase A (Kata) activity and  
RL resistance to hydrogen peroxide in Pseudomonas aeruginosa.";  
RN J. Bacteriol. 181:3730-3742(1999).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 15692 / PA01;  
RX MEDLINE=20437337; PubMed=10984043;  
RA Stover C.K., Pham X.-O.T., Erwin A.L., Mizoguchi S.D., Warren P.,  
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,  
RA Garber R.L., Collier L., Tolerino E., Westbrook-Wadman S., Yuan Y.,  
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Laidig K., Lam R.M.,  
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,  
RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;  
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an  
RT opportunistic pathogen.";  
RL Nature 406:959-964(2000).  
CC -1- FUNCTION: DECOMPOSES HYDROGEN PEROXIDE IN WATER AND OXYGEN; SERVES  
CC TO PROTECT CELLS FROM THE TOXIC EFFECTS OF HYDROGEN PEROXIDE.  
CC -1- CATALYTIC ACTIVITY: 2 H(2)O(2) = O(2) + 2 H(2)O.  
CC -1- COFACTOR: HEME GROUP.  
CC -1- ENZYME REGULATION: BY PEROXIDE AND BFR-BOUND IRON.  
CC -1- SUBUNIT: HETEROMULTIMER, POSSIBLY AN ALPHA(2)BETA-HETEROTRIMER  
CC WHERE THE ALPHA SUBUNIT IS A 56 KDA PROTEIN AND THE BETA SUBUNIT A  
CC 45 KDA PROTEIN.  
CC -1- SIMILARITY: BELONGS TO THE CATALASE FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
CC EMBL: AF047025; AAC03118.1; -  
CC EMBL: AE004841; AAC07624.1; -  
CC InterPro: IPR002226; Catalase.

DR Pfam: PF00199; catalase; 1.  
 DR PRINTS: PR00067; CATALASE.  
 DR Prodom: PD000510; CATALASE; 1.  
 DR PROSITE: PS00437; CATALASE\_1; 1.  
 DR PROSITE: PS00438; CATALASE\_2; 1.  
 KW Oxidoreductase; Peroxidase; Iron; Heme; Hydrogen peroxide;  
 Complete proteome.  
 FT AC1\_SITE 55 BY SIMILARITY.  
 FT ACT\_SITE 128 BY SIMILARITY.  
 FT BINDING 338 PROXIMAL HEME LIGAND (BY SIMILARITY).  
 SQ SEQUENCE 482 AA; 55589 MW; 84E5ADA647CAB414 CRC64;

Query Match 77.8%; Score 63; DB 1; Length 482;  
 Best Local Similarity 82.4%; Pred. No. 0.0022;  
 Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

2 EEKPLTTAAAXAPVYN 18  
 11111111111111111111  
 2 EEKRLTTAAGAPVVDN 18

RESULT 2  
 CATA\_PROMI STANDARD; PRT; 484 AA.  
 ID CATA\_PROMI  
 AC P42321:  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE CATALASE (EC 1.11.1.6).  
 GN KATA.  
 OS Proteus mirabilis.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Proteus.  
 NCBI\_TaxID=584;  
 RN [1]  
 RP SEQUENCE AND SEQUENCE OF 1-305 FROM N.A.  
 RC STRAIN=PR;  
 RX MEDLINE=95305957; PubMed=7786407;  
 RA Bazy A., Bracchi V., Sterjades R., Chroboczek J., Thibault P.,  
 RA Gagnon J., Jouve H.-M., Hudry-Clergeon G.;  
 RT "Complete amino acid sequence of Proteus mirabilis PR catalase,  
 RT occurrence of a methionine sulfone in the close proximity of the  
 RT active site.";  
 RT J. Protein Chem. 14:59-72(1995).  
 [2]  
 X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS).  
 RC STRAIN=PR;  
 RX MEDLINE=95311317; PubMed=7791219;  
 RA Gouet P., Jouve H.-M., Dideberg O.;  
 RT "Crystal structure of Proteus mirabilis PR catalase with and without  
 RT bound NADPH.";  
 RT J. Mol. Biol. 249:933-954(1995).  
 CC -1- FUNCTION: DECOMPOSES HYDROGEN PEROXIDE IN WATER AND OXYGEN. SERVES  
 CC TO PROTECT CELLS FROM THE TOXIC EFFECTS OF HYDROGEN PEROXIDE.  
 CC -1- CATALYTIC ACTIVITY: 2 H(2)O(2) = O(2) + 2 H(2)O.  
 CC -1- COFACTOR: HEME GROUP AND NADP.  
 CC -1- SUBUNIT: HOMOTETRAMER.  
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.  
 CC -1- MASS SPECTROMETRY: MW=55643; MW\_ERR=5; METHOD=ELECTROSPRAY.  
 CC -1- SIMILARITY: BELONGS TO THE CATALASE FAMILY.  
 DR PDB: 2CAF; 08-DEC-96.  
 DR PDB: 2CAF; 07-DEC-96.  
 DR PDB: 2CAG; 07-DEC-96.  
 DR PDB: 2CAH; 11-JAN-97.  
 DR InterPro: IPR002226; Catalase.  
 DR Pfam: PF00199; catalase; 1.  
 DR PRINTS: PR00067; CATALASE.  
 DR Prodom: PD000510; CATALASE; 1.  
 DR PROSITE: PS00437; CATALASE\_1; 1.  
 DR PROSITE: PS00438; CATALASE\_2; 1.  
 KW Oxidoreductase; Peroxidase; Iron; Heme; Hydrogen peroxide; NADP;  
 3d-structure.

FT MOD\_RES 53 53 METHIONINE SULFONE.  
 FT ACT\_SITE 54 54  
 FT ACT\_SITE 127 127  
 FT BINDING 337 337  
 SQ SEQUENCE 484 AA; 55614 MW; ADC25F3CBA1F5C50 CRC64;

Query Match 65.4%; Score 53; DB 1; Length 484;  
 Best Local Similarity 75.0%; Pred. No. 0.11;  
 Matches 12; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 EEKPLTTAAAXAPVYN 18  
 11111111111111111111  
 DB 2 EEKRLTTAAGAPVVDN 17

RESULT 3  
 RL7\_STRAT STANDARD; PRT; 128 AA.  
 ID RL7\_STRAT  
 AC P29342;  
 DT 01-DEC-1992 (Rel. 24, Created)  
 DT 01-DEC-1992 (Rel. 24, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE 50S RIBOSOMAL PROTEIN L7/L12.  
 GN RPLL.  
 OS Streptomyces antibioticus.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.  
 NCBI\_TaxID=1890;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92380478; PubMed=1511874;  
 RA Parra F., Blanco G., Alonso J.M., Balbin M., Mendez C., Salas J.A.;  
 RT "Cloning and sequence of a gene encoding the L7/L12 ribosomal protein  
 RT equivalent of Streptomyces antibioticus.";  
 RL Gene 118:127-129(1992).  
 CC -1- FUNCTION: SEEMS TO BE THE BINDING SITE FOR SEVERAL OF THE FACTORS  
 CC INVOLVED IN PROTEIN SYNTHESIS AND APPEARS TO BE ESSENTIAL FOR  
 CC ACCURATE TRANSLATION (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE L12P FAMILY OF RIBOSOMAL PROTEINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-slb.ch/announce/>  
 CC or send an email to [license@isb-slb.ch](mailto:license@isb-slb.ch)).  
 CC -----  
 DR EMBL: M89911; AAA26811.1; -  
 DR PIR: JC1273; JC1273.  
 DR HSSP: P02392; ICTF.  
 DR InterPro: IPR000206; Ribosomal\_L12.  
 DR Pfam: PF00542; Ribosomal\_L12; 1.  
 DR Prodom: PD001326; Ribosomal\_L12; 1.  
 KW Ribosomal protein.  
 SQ SEQUENCE 128 AA; 13272 MW; F5C3EE4F45D06068 CRC64;

Query Match 55.6%; Score 45; DB 1; Length 128;  
 Best Local Similarity 66.7%; Pred. No. 0.64;  
 Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 2 EEKPLTTAAAXAPV 16  
 11111111111111111111  
 DB 28 EEKPLTTAAAXAPV 42

RESULT 4  
 BCA\_STRVL STANDARD; PRT; 483 AA.  
 ID BCA\_STRVL  
 AC P33569;  
 DT 01-FEB-1994 (Rel. 28, Created)

DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 01-OCT-1996 (Rel. 34, Last annotation update)  
 DE BROMOPEROXIDASE-CATALASE (EC 1.11.1.1.).  
 GN BCA.  
 OS Streptomyces violaceus (Streptomyces venezuelae).  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae.  
 CC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.  
 CX NCBI\_TaxID=1936;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 10712 / ISP5230;  
 RX MEDLINE=97022081; Pubmed=8868441;  
 RA Pacey S., Gross F., Vining L.C., Yang K., van Pee K.-H.;  
 RT "Cloning, sequencing and disruption of a bromoperoxidase-catalase  
 gene in Streptomyces venezuelae: evidence that it is not required for  
 chlorination in chloramphenicol biosynthesis.";  
 RT Microbiology 142:657-665(1996).  
 RL -1- CATALYTIC ACTIVITY: 2 H(2)O(2) - O(2) + 2 H(2)O.  
 CC -1- SIMILARITY: BELONGS TO THE CATALASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: X74791; CAA52796.1; -  
 CC PIR: S37055; S37055.  
 DR HSSP: P42321; ZCAH.  
 DR InterPro: IPR002226; Catalase.  
 DR Pfam: PF00199; catalase; 1.  
 DR PRINTS: PR00067; CATALASE.  
 DR ProDom: PD000510; Catalase; 1.  
 DR PROSITE: PS00437; CATALASE\_1; 1.  
 DR PROSITE: PS00438; CATALASE\_2; 1.  
 DR Oxidoreductase; Peroxidase; Iron; Heme; Hydrogen peroxide.  
 KM ACT\_SITE 54 54  
 FT ACT\_SITE 127 127 BY SIMILARITY.  
 FT BINDING 337 337 PROXIMAL HEME LIGAND (BY SIMILARITY).  
 FT SEQUENCE 483 AA; 54087 MW; B58CF8230B8A2F55 CRC64;  
 SQ  
 Query Match 53.1%; Score 43; DB 1; Length 483;  
 Best Local Similarity 69.2%; Pred. No. 5.4;  
 Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

CC -1- CATALYTIC ACTIVITY: 2 H(2)O(2) - O(2) + 2 H(2)O.  
 CC -1- COFACTOR: HEME GROUP.  
 CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE CATALASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: U07800; AAA18481.1; -  
 CC HSSP: P42321; ZCAH.  
 DR InterPro: IPR002226; Catalase.  
 DR Pfam: PF00199; catalase; 1.  
 DR PRINTS: PR00067; CATALASE.  
 DR ProDom: PD000510; Catalase; 1.  
 DR PROSITE: PS00437; CATALASE\_1; 1.  
 DR PROSITE: PS00438; CATALASE\_2; 1.  
 DR Oxidoreductase; Peroxidase; Iron; Heme; Hydrogen peroxide.  
 KM ACT\_SITE 57 57  
 FT ACT\_SITE 130 130 BY SIMILARITY.  
 FT BINDING 340 340 PROXIMAL HEME LIGAND (BY SIMILARITY).  
 FT SEQUENCE 482 AA; 54508 MW; 7CB73E08975C219F CRC64;  
 SQ

Query Match 50.6%; Score 41; DB 1; Length 482;  
 Best Local Similarity 75.0%; Pred. No. 12;  
 Matches 9; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 7 LTTAAXAPVYXN 18  
 DB 9 LTTAAGAPVADN 20  
 RESULT 6  
 ACML\_DROME STANDARD; PRT: 722 AA.  
 ID ACML\_DROME  
 AC P16395;  
 DT 01-AUG-1990 (Rel. 15, Last sequence update)  
 DT 01-AUG-1990 (Rel. 15, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE MUSCARINIC ACETYLCHOLINE RECEPTOR DML.  
 GN ACRC OR MACR-60C.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 CX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=OREGON-R;  
 RX MEDLINE=90046926; Pubmed=2510174;  
 RA Shapiro R.A., Wakimoto B.T., Subers E.M., Nathanson N.M.,  
 RT "Characterization and functional expression in mammalian cells of  
 genomic and cDNA clones encoding a Drosophila muscarinic  
 acetylcholine receptor.";  
 RT Proc. Natl. Acad. Sci. U.S.A. 86:9039-9043(1989).  
 RL [2]  
 RN SEQUENCE FROM N.A.  
 RP MEDLINE=9005981; Pubmed=2507354;  
 RA Onal T., Fitzgerald M.G., Arakawa S., Gocayne J.D., Urquhart D.A.,  
 RA Hall L.M., Fraser C.M., McCombie W.R., Venter J.C.;  
 RT "Cloning, sequence analysis and chromosome localization of a  
 Drosophila muscarinic acetylcholine receptor.";  
 RL FEBS Lett. 255:219-225(1989).  
 CC -1- FUNCTION: THE MUSCARINIC ACETYLCHOLINE RECEPTOR MEDIATES VARIOUS  
 CELLULAR RESPONSES, INCLUDING INHIBITION OF ADENYLATE CYCLASE,  
 BREAKDOWN OF PHOSPHOINOSITIDES & MODULATION OF POTASSIUM CHANNELS  
 THROUGH THE ACTION OF G PROTEINS. PRIMARY TRANSDUCING EFFECT IS  
 PI TURNOVER.  
 CC

CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.  
 CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL: M27495; AA85449.1; -  
 DR EMBL: M23412; AAA28676.1; ALT\_INIT.  
 DR PIR: A36191; A36191.  
 DR PIR: S05661; S05661.  
 DR GCRDB: GCR\_0024; -  
 DR GCRDB: GCR\_0025; -  
 DR FlyBase: FBgn0000037; macr-60C.  
 DR InterPro: IPR000276; GPCR\_Rhodopsn.  
 DR Pfam: PF00001; 7tm\_1; 1.  
 DR PRINTS: PR00237; GPCRHHODPSN.  
 DR PRINTS: PR00243; MUSCARINICR.  
 DR PROSITE: PS00237; G-PROTEIN RECP\_FL1; 1.  
 DR PROSITE: PS50262; G-PROTEIN RECP\_FL2; 1.  
 KM Postsynaptic membrane: Ionic channel; Glycoprotein; Transmembrane;  
 KM Phosphorylation; Multigene family; G-protein coupled receptor.  
 FT DOMAIN 1 26  
 FT TRANSMEM 27 49  
 FT TRANSMEM 50 60  
 FT TRANSMEM 61 81  
 FT TRANSMEM 82 100  
 FT TRANSMEM 101 120  
 FT TRANSMEM 121 140  
 FT TRANSMEM 141 162  
 FT TRANSMEM 163 184  
 FT TRANSMEM 185 208  
 FT TRANSMEM 209 634  
 FT TRANSMEM 635 656  
 FT TRANSMEM 657 672  
 FT TRANSMEM 673 693  
 FT DOMAIN 694 722  
 FT CARBOHYD 4 4  
 FT CARBOHYD 7 7  
 FT CONFLICT 75 75  
 FT CONFLICT 119 119  
 FT CONFLICT 136 136  
 FT CONFLICT 147 147  
 FT CONFLICT 250 250  
 FT CONFLICT 346 362  
 FT CONFLICT 381 381  
 FT CONFLICT 451 451  
 FT CONFLICT 451 451  
 FT CONFLICT 606 606  
 FT CONFLICT 614 614  
 FT CONFLICT 654 654  
 FT CONFLICT 688 688  
 FT CONFLICT 714 722  
 FT SEQUENCE 722 AA; 78237 MW; FB35417CB2202A61 CRC64;  
 Query Match 49.4%; Score 40; DB 1; Length 722;  
 Best Local Similarity 52.9%; Pred. No. 26;  
 Matches 9; Conservative 1; Mismatches 7; Indels 0; Gaps 0;  
 QY 3 EKTPLTTAAKAPVYXNA 19  
 DB 538 EDGPTTAAAPLASAA 554  
 RESULT 7  
 H2B\_AGABI STANDARD; PRT; 143 AA.  
 AC P78567;  
 DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE HISTONE H2B.  
 GN HTBA.  
 OS Agaricus bisporus (Common mushroom).  
 OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;  
 OC Agaricales; Agaricaceae; Agaricus.  
 OX NCBI\_TaxID=5341;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=HORSF 01;  
 RX MEDLINE=97111990; PubMed=8953726;  
 RA Sonnenberg A.S.M., de Groot P.W.J., Schaap P.J., Baars J.J.P.,  
 RA Visser J., van Gilsen L.J.L.D.;  
 RT Isolation of expressed sequence tags of Agaricus bisporus and their  
 RT assignment to chromosomes.\*  
 RL Appl. Environ. Microbiol. 62:4542-4547(1996).  
 CC -1- SUBUNIT: THE NUCLEOSOME IS AN OCTAMER CONTAINING TWO MOLECULES OF  
 CC H2A, H2B, H3, AND H4; WHICH WRAP APPROXIMATELY 146 BP OF DNA.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -1- SIMILARITY: BELONGS TO THE HISTONE H2B FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL: X94188; CA63898.1; -  
 DR InterPro: IPR000558; Histone\_H2B.  
 DR InterPro: IPR000166; Histone\_core.  
 DR Pfam: PF00125; Histone; 1.  
 DR PRINTS: PR00621; HISTONEH2B.  
 DR ProDom: PD000497; Histone\_H2B.  
 DR SMART: SM00427; H2B; 1.  
 DR PROSITE: PS00357; HISTONE\_H2B; 1.  
 KM Chromosomal protein; Nucleosome core; Nuclear protein; DNA-binding.  
 KM SEQUENCE 143 AA; 15165 MW; 69640102E31F4B56 CRC64;  
 Query Match 48.1%; Score 39; DB 1; Length 143;  
 Best Local Similarity 50.0%; Pred. No. 7.4;  
 Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;  
 QY 4 KTEPLTTAAKAPVYXNA 19  
 DB 11 KAPASTASKAPVKSDA 26  
 RESULT 8  
 DMPC\_PSESP STANDARD; PRT; 486 AA.  
 ID DMPC\_PSESP  
 AC P19059;  
 DT 01-NOV-1990 (Rel. 16, Created)  
 DT 01-NOV-1990 (Rel. 16, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE 2-HYDROXYMUCONIC SEMIALDEHYDE DEHYDROGENASE (EC 1.2.1.-) (HMSD).  
 GN DMPC.  
 OS Pseudomonas sp. (strain CF600).  
 OC Plasmid pVil150.  
 OC Bacteria; Proteobacteria.  
 OX NCBI\_TaxID=306;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=90304229; PubMed=2194577;  
 RA Nordlund I., Shingler V.;  
 RT Nucleotide sequences of the meta-cleavage pathway enzymes 2-  
 RT hydroxymucronic semialdehyde dehydrogenase and 2-hydroxymucronic  
 RT semialdehyde hydrolase from Pseudomonas CF600.\*;  
 RL Biochim. Biophys. Acta 1049:227-230(1990).  
 CC -1- FUNCTION: 2-HYDROXYMUCONIC ACID SEMIALDEHYDE CAN BE CONVERTED TO

2-HYDROXYMUTANT-2,4-DIENATE EITHER DIRECTLY BY THE ACTION OF  
THREE SEQUENTIAL ENZYMES, THE FIRST OF WHICH IS HMD. -  
-1- PATHWAY: METACLEAVAGE PATHWAY FOR THE DEGRADATION OF PHENOLS,  
-1- CRESOLS AND CATECHOL. PHENOL METABOLISM.  
-1- SIMILARITY: BELONGS TO THE ALDEHYDE DEHYDROGENASES FAMILY.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL: X52805; CNA36992.1; -  
PIR: S10772; S10772.  
DR HSSP: P20000; 1A4.  
DR InterPro: IPR002086; Aldehyde\_dehydr.  
DR Pfam: PF00171; Aldehyd. 1.  
DR PROSITE: PS00070; ALDEHYDE\_DEHYDR\_CYS; 1.  
DR PROSITE: PS00687; ALDEHYDE\_DEHYDR\_GLU; 1.  
DR Aromatic hydrocarbons catabolism: Oxidoreductase; NAD; Plasmid.  
ACT SITE 254 254 BY SIMILARITY.  
ACT SITE 288 288 BY SIMILARITY.  
SEQUENCE 486 AA; 51682 MW; F906FCA64185AA68 CRC64;  
-----  
Query Match 48.1%; Score 39; DB 1; Length 486;  
Best Local Similarity 52.9%; Pred. No. 26;  
Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0;  
-----  
OY 3 EKPPLTTAAAXAPVYXNA 19  
DB 181 EEPPLTALLCEVMOAA 197  
-----  
RESULT 9  
YB64\_YEAST STANDARD: PRT: 527 AA.  
ID YB64\_YEAST  
AC 01-OCT-1994 (Rel. 30, Created)  
DT 01-OCT-1994 (Rel. 30, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last annotation update)  
DE HYPOHETICAL 57.2 KDA PROTEIN IN MET8-HPC2 INTERGENIC REGION.  
GN YBR214W OR YBR1501  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Saccharomycetes: Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetaceae; Saccharomycetaceae; Saccharomycetes.  
NCBI\_TaxID=4932;  
OX 11  
SEQUENCE FROM N.A.  
STRAIN=S288C;  
Rieger M.;  
Submitted (AUG-1994) to the EMBL/Genbank/DBJ databases.  
-----  
[12]  
RN SEQUENCE OF 114-527 FROM N.A.  
RP SPRAIN-S288C;  
RC Dubois E., el Bakoury M., Glandsdorff N., Mesenguy F., Plerard A.,  
RA Scherens B., Vlerendeels F.;  
RL Submitted (AUG-1994) to the EMBL/Genbank/DBJ databases.  
CC -1- SIMILARITY: TO YEAST YGL056C AND S.POMBE SDS23.  
CC -1- SIMILARITY: CONTAINS 2 CBS DOMAINS.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL: Z36083; CNA85178.1; -

DR PIR: S46088; S46088.  
DR SGD: S0000418; SD524.  
DR InterPro: IPR000644; CBS.  
DR Pfam: PF00571; CBS; 3.  
DR SMART: SM00116; CBS; 2.  
DR KMW Hypothetical protein: Repeat; CBS domain.  
FT DOMAIN 196 251 CBS 1.  
FT DOMAIN 283 335 CBS 2.  
SQ SEQUENCE 527 AA; 57187 MW; DC2741550A69C154 CRC64;  
-----  
Query Match 48.1%; Score 39; DB 1; Length 527;  
Best Local Similarity 61.5%; Pred. No. 28;  
Matches 8; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
-----  
OY 2 EKPPLTTAAAXAP 14  
DB 52 EEPPLTAAAP 64  
-----  
RESULT 10  
VNQC\_INBLE STANDARD: PRT: 560 AA.  
ID VNQC\_INBLE  
AC P04665;  
DT 13-AUG-1987 (Rel. 05, Created)  
DT 13-AUG-1987 (Rel. 05, Last sequence update)  
DT 01-MAY-1992 (Rel. 22, Last annotation update)  
DE NUCLEOPROTEIN.  
GN NP Influenza B virus (strain B/Lee/40).  
OS Influenza B virus (strain B/Lee/40).  
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;  
OC Influenza virus A and B group; Influenza B virus.  
OX NCBI\_TaxID=11535;  
RN 11  
RP SEQUENCE FROM N.A.  
RX MEDLINE=84174071; PubMed=6324462;  
RA Briedis D.J., Tobin M.;  
RT Influenza B virus genome: complete nucleotide sequence of the  
RT Influenza B/Lee/40 virus genome RNA segment 5 encoding the  
RT nucleoprotein.  
RT nucleoprotein.  
RL Virology 133:448-455(1984).  
CC -1- FUNCTION: ENCAPSIDATES THE NEGATIVE STRAND VIRAL RNA.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL: K01395; AAA3689.1; -  
DR InterPro: IPR002141; FLU\_NP.  
DR Pfam: PF00506; FLU\_NP; 1.  
DR Nucleoprotein.  
KW SEQUENCE 560 AA; 61770 MW; 74CAFEAF9E75A695 CRC64;  
-----  
Query Match 48.1%; Score 39; DB 1; Length 560;  
Best Local Similarity 44.4%; Pred. No. 30;  
Matches 8; Conservative 3; Mismatches 7; Indels 0; Gaps 0;  
-----  
OY 2 EKPPLTTAAAXAPVYXNA 19  
DB 19 EEPPLTSGATPPIIKPA 36  
-----  
RESULT 11  
ID GYRB\_MYXXA STANDARD: PRT: 815 AA.  
AC GYRB\_MYXXA  
DT 15-DEC-1998 (Rel. 37, Created)

2002085FBB805B32 CRC64;

	48.1%;	Score 39;	DB 1;	Length 815;
Query Match				
Best Local Similarity	66.7%;	Pred. No. 44;		
Matches	8;	Conservative 1;	Mismatches 3;	Indels 0;
				Gaps 0
0Y	3 EKKPLTTAAAP	14		
		1		
		1		
Db	2 EKKPAGSAAP	13		

RESULT	12		
GBP_PSESE			
ID		STANDARD:	PRT: 143 AA.
AC	027913,		
DT	15-JUL-1999 (Rel. 38, Created)		
DT	15-JUL-1999 (Rel. 38, Last sequence update)		
DE	15-JUL-1999 (Rel. 38, Last annotation update)		
OS	GROWTH-BLOCKING PEPTIDE PECCURSOR (GBP).		
OS	Pseudaleitia separata (Armyworm).		
OC	Eukaryota: Metazoa: Arthropoda: Tracheata: Hexapoda: Insecta;		
OC	Pterygota: Neoptera: Endopterygota: Lepidoptera: Glossata; Dityr		
OC	Noctuoidea: Noctuidae: Hadeninae: Pseudaleitia.		
OX	NCBI_TaxId=7105;		

```

RN [1]
RN SEQUENCE FROM N.A.
RX MEDLINE=96103377; PubMed=7496538;
RA Hayakawa Y., Ohnishi A., Yamamaka A., Izumi S., Tomino S.;
RT "Molecular cloning and characterization of cDNA for insect biogenic
RT peptide, growth-blocking peptide.",
RL Febs Lett. 376:185-189(1995).
RN [2]
RN SEQUENCE FROM N.A.
RX MEDLINE=98316655; PubMed=9654083;
RA Hayakawa Y., Noguchi H.;
RT "Growth-blocking peptide expressed in the insect nervous system:
RT cloning and functional characterization.",
RL Eur. J. Biochem. 253:810-816(1998).
RN [3]
RN STRUCTURE BY NMR OF 121-143.
RX TISSUE=Hemolymph;
RX MEDLINE=99107831; PubMed=9890941;
RA Alzawa T., Fujitani N., Hayakawa Y., Ohnishi A., Okubo T., Kumaki Y.,
RA Kawanu K., Hikichi K., Nitta K.;
RT "Solution structure of an insect growth factor, growth-blocking
RT peptide.";
RL J. Biol. Chem. 274:1887-1890(1999).
CC -1- FUNCTION: BIOGENIC PEPTIDE THAT PREVENTS, IN LEPTODOPTERAN, THE
CC ONSET OF METAMORPHOSIS FROM LARVA TO PUPA. THIS GROWTH-BLOCKING
CC PEPTIDE HAS REPRESSIVE ACTIVITY AGAINST JUVENILE HORMONE ESTERASE.
CC -1- SIMILARITY: BELONGS TO THE GBP / PSP1 / PARALYTIC PEPTIDE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; S80564; AAB3742.1;
DR EMBL; AB012294; BA332793.1;
DR InterPro: IPR003463; GBP_PSP.
DR Pfam: PF02425; GBP_PSP; 1.
KW Hemolymph; Signal; 3D-structure.
FT SIGNAL 1 ?
FT PROPEP 120 ?
FT PEPTIDE 121 143 POTENTIAL.
FT DISULFIDE 127 139 GROWTH-BLOCKING PEPTIDE.
SQ SEQUENCE 143 AA; 15256 MW; 15256 MW; DEBPS27956840EB CRC64;

```

0000002/300840EB CRC64;

Query Match	46.9%	Score 38;	DB 1;	Length 143;
Best Local Similarity	50.0%;	Pred. NO. 11;		
Matches	10; Conservative	1; Mismatches	7; Indels	2; Gaps
QY	2 EERTP-LTTAAAPVYXNA 19			
	:			
Db	83 DEVTPTATTTTAAAPVYXNA 102			

RESULT	13
PLD_CORPS	
1D	PLD_CORPS
AC	P20626; Q59314; STANDARD; PRT; 307 AA.
DT	01-FEB-1991 (Rel. 17, Created)
DT	01-FEB-1991 (Rel. 17, Last sequence update)
DT	01-NOV-1997 (Rel. 35, Last annotation update)
DE	PHOSPHOLIPASE D PRECURSOR (EC 3.1.4.4) (PLD) (CHOLINE PHOSPHATASE) PLD.
GN	Corynebacterium pseudotuberculosis.
OS	Bacteria; Firmicutes; Actinobacteria; Actinobacteriales; Actinobacteriaceae; Corynebacteriinae; Corynebacteriaceae; Corynebacterium.
OC	Corynebacterium.
CC	NCBI_TaxID=1719;
XX	[1]
NN	





RX MEDLINE=98332770; PubMed=9665876;  
 RA Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.C.,  
 RA Dodson R., Gwin M., Hickey E.K., Clayton R., Ketchum K.A.,  
 RA Sodergren E., Haidam J.M., McLeod M.P., Salzberg S., Peterson J.,  
 RA Khalek H., Richardson D., Howell J.K., Chidambaram M., Uterback T.,  
 RA McDonald L., Attiach P., Bowman C., Cotton M.D., Fujii C., Garland S.,  
 RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,  
 RA Venter J.C.;  
 RT "Complete genome sequence of Treponema pallidum, the syphilis  
 RT spirochete.";  
 RL Science 281:375-388(1998).  
 CC -!- FUNCTION: THIS PROTEIN PROMOTES THE GTP-DEPENDENT TRANSLLOCATION  
 CC OF THE NASCENT PROTEIN CHAIN FROM THE A-SITE TO THE P-SITE OF THE  
 CC RIBOSOME.  
 CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.  
 CC -!- SIMILARITY: BELONGS TO THE GTP-BINDING ELONGATION FACTOR FAMILY.  
 CC EF-G/EF-2 SUPERFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: AE001248; AAC65735.1; -  
 DR TIGR: TP0767; -  
 DR InterPro: IPR000640; EFG\_C.  
 DR InterPro: IPR000795; GTP\_EFTU.  
 DR Pfam: PF00679; EFG\_C.1.  
 DR Pfam: PF00009; GTP\_EFTU; 1.  
 DR PRINTS: PR00315; ELONGATNFCT.  
 DR PROSITE: PS00301; EFATOR\_GTP; 1.  
 DR Elongation factor; Protein biosynthesis; GTP-binding;  
 KW Multigene family; Complete proteome.  
 FT NP\_BIND 15 22 GTP (BY SIMILARITY).  
 FT NP\_BIND 82 86 GTP (BY SIMILARITY).  
 FT NP\_BIND 136 139 GTP (BY SIMILARITY).  
 FT NP\_BIND 139 139 GTP (BY SIMILARITY).  
 SQ SEQUENCE 695 AA; 76831 MW; 46529989BFE97E4E CRC64;

Query Match 46.9%; Score 38; DB 1; Length 695;  
 Best Local Similarity 60.0%; Pred. No. 55;  
 Matches 9; Conservative 1; Mismatches 5; Indels 0; Gaps 0;  
 2 EEXPLTTAAXAPV 16  
 11 1111111111  
 Db 295 EEPVTLSTDADAPV 309

Search completed: February 28, 2002, 11:39:52  
 Job time: 308 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 28, 2002, 11:39:32 : Search time 35.84 seconds  
(without alignments)  
77.544 Million cell updates/sec

Title: US-09-359-426C-2  
Perfect score: 81  
Sequence: 1 XEKKTP/TTAAKAPVXNA 19

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: SP-archaea:\*  
2: SP-bacteria:\*  
3: SP-fungi:\*  
4: SP-human:\*  
5: SP-invertebrate:\*  
6: SP-mammal:\*  
7: SP-mmc:\*  
8: SP-organelle:\*  
9: SP-phage:\*  
10: SP-plant:\*  
11: SP-protent:\*  
12: SP-virus:\*  
13: SP-vertebrate:\*  
14: SP-unclassified:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	46	56.8	458	5	076672
2	45	55.6	484	5	09RG14
3	44	54.3	314	2	09A119
4	44	54.3	427	10	023188
5	43	53.1	506	2	P77924
6	43	53.1	1240	12	Q9DMH8
7	42	51.9	210	2	Q9RN02
8	42	51.9	211	2	Q9X6X8
9	42	51.9	436	10	Q9RT45
10	42	51.9	487	2	Q9RJK9
11	41	50.6	132	10	Q9SR04
12	41	50.6	417	2	Q34272
13	41	50.6	1012	5	Q9V745
14	40	49.4	171	5	Q9U3E3
15	40	49.4	212	5	Q9U4Y3
16	40	49.4	252	1	Q9Y984
17	40	49.4	394	2	Q9A4B0
18	40	49.4	477	2	Q9RD97
19	40	49.4	498	12	Q79665

20	40	49.4	594	5	Q9BLH5	Q9BLH5 halocynthia
21	40	49.4	788	5	Q9W180	Q9W180 drosophila
22	40	49.4	846	13	057577	057577 cynops pyrr
23	40	49.4	1360	5	002006	002006 drosophila
24	40	49.4	1473	5	Q9VHP9	Q9VHP9 drosophila
25	39.5	48.8	405	10	Q03462	Q03462 zea mays (m
26	39.5	48.8	1793	5	Q9W596	Q9W596 drosophila
27	39	48.1	255	10	Q9LGM5	Q9LGM5 oryza sativ
28	39	48.1	357	5	Q9GV02	Q9GV02 leishmania
29	39	48.1	410	2	Q9PD09	Q9PD09 streptococ
30	39	48.1	484	2	Q9CPK5	Q9CPK5 pasteurella
31	39	48.1	484	12	Q9WA02	Q9WA02 human adeno
32	39	48.1	507	2	Q9A8Y1	Q9A8Y1 caulobacter
33	39	48.1	518	13	Q9PTY0	Q9PTY0 cyrilus ca
34	39	48.1	572	10	Q9LJG2	Q9LJG2 arabidopsis
35	39	48.1	952	12	Q99174	Q99174 human adeno
36	39	48.1	1353	5	Q18182	Q18182 caenorhabd
37	39	48.1	4900	5	Q9N541	Q9N541 caenorhabd
38	38.5	47.5	842	5	Q9VJH6	Q9VJH6 drosophila
39	38.5	47.5	1658	2	Q9RM11	Q9RM11 delinococcus
40	38	46.9	167	2	Q31102	Q31102 rhodospirill
41	38	46.9	195	2	Q915W8	Q915W8 streptococ
42	38	46.9	216	2	Q9KFA0	Q9KFA0 streptococ
43	38	46.9	224	2	Q9RFJ3	Q9RFJ3 streptococ
44	38	46.9	225	5	Q9VE40	Q9VE40 drosophila
45	38	46.9	229	4	Q9H898	Q9H898 homo sapien

## ALIGNMENTS

RESULT 1  
ID 076672 PRELIMINARY: PRT: 458 AA.  
AC 076672;  
DT 01-NOV-1998 (TREMBLrel. 08, Created)  
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE HYPOTHETICAL PROTEIN H34124.2.  
GN H34124.2.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
OC Rhabditidae; Pelodierinae; Caenorhabditis.  
OX NCBI\_Taxid=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RX MEDLINE=99069613; PubMed=9851916;  
RA None;  
RT "Genome sequence of the nematode C. elegans: a platform for  
investigating biology. The C. elegans Sequencing Consortium.";  
RL Science 282:2012-2018(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RA Latreille P., Wamsley P., O'Brien D.;  
RT "The sequence of C. elegans cosmid H34124.";  
RL Submitted (JUL-1998) to the EMBL/Genbank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RA Waterston R.;  
RT Submitted (MAR-2001) to the EMBL/Genbank/DBJ databases.  
DR EMBL: AF078784; AKI8967.1; -; ZAF5BC54EDA0C324 CRC64;  
SQ SEQUENCE 458 AA: 47285 MW: 2AF5BC54EDA0C324 CRC64;

Query Match 56.8%; Score 46; DB 5; Length 458;  
Best local Similarity 56.2%; Pred. No. 4.5;  
Matches 9; Conservative 3; Mismatches 4; Indels 0; Gaps 0;  
QY 4 KTEPLTTAAKAPVXNA 19  
|||||:|:|:|

Db 386 KPTLTSSSARVINNA 401

RESULT 2

Q9RG14 PRELIMINARY; PRT; 484 AA.

AC Q9RG14; 01-MAY-2000 (TREMblrel. 13, Created)

DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)

DE 01-JUN-2001 (TREMblrel. 17, Last annotation update)

DE CATALASE (EC 1.11.1.6).

GN KATA.

OS Actinobacillus actinomycetemcomitans (Haemophilus actinomycetemcomitans).

OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae; Actinobacillus.

NCBI\_TaxID=714;

[1]

SEQUENCE FROM N.A.

RP STRAIN=YANAL;

RC MEDLINE=20042351; PubMed=10572134;

RA Thomson V.J., Bhattacharjee M.K., Fine D.H., Derbyshire K.M., Figurski D.H.;

RT "Direct Selection of IS903 Transposon Insertions by Use of a Broad-Host-Range Vector: Isolation of Catalase-Deficient Mutants of Actinobacillus actinomycetemcomitans."

RL J. Bacteriol. 181:7298-7307(1999).

CC -1- FUNCTION: OCCURS IN ALMOST ALL AEROBICALLY RESPIRING ORGANISMS AND SERVES TO PROTECT CELLS FROM THE TOXIC EFFECTS OF HYDROGEN PEROXIDE (BY SIMILARITY).

CC -1- CATALYTIC ACTIVITY: 2 H(2)O(2) = O(2) + 2 H(2)O.

CC EMBL: AF162654; AAF17882.1; -

DR HSSP: P42321; 2CAE.

DR InterPro: IPR002226; Catalase.

DR Pfam: PF00199; catalase; 1.

DR PRINTS: PR00067; CATALASE.

DR PRODOM: PD000510; Catalase; 1.

DR PROSITE: PS00437; CATALASE.1; 1.

DR PROSITE: PS00438; CATALASE.2; 1.

KW Heme: Hydrogen peroxide: Iron: Oxidoreductase: Peroxidase.

SO SEQUENCE 484 AA; 54961 MW; D0E523AB2557D8CB CRC64;

Query Match 55.6%; Score 45; DB 2; Length 484; Best Local Similarity: 83.3%; Pred. No. 7.1; Matches 10; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 7 LTTAAAPVYXN 18  
|||||

Db 14 LTTAAGAPVDN 25

RESULT 3

O9A119 PRELIMINARY; PRT; 314 AA.

AC O9A119; 01-JUN-2001 (TREMblrel. 17, Created)

DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)

DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)

DE HYPOTHEICAL 34.4 KDA PROTEIN.

OS Escherichia coli.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.

NCBI\_TaxID=562;

[1]

SEQUENCE FROM N.A.

RP STRAIN=AL862;

RC MEDLINE=21101044; PubMed=11159989;

RA Lalioui L., Le Bouguenec C.C.;

RT "afa-8 Gene cluster is carried by a pathogenicity island inserted into the tRNA(Phe) of human and bovine pathogenic Escherichia coli isolates."

RL Infect. Immun. 69:937-948(2001).

DR EMBL: AF286671; AAK27331.1; -

KW Hypothetical protein.

SO SEQUENCE 314 AA; 34415 MW; 2D5F7342C817E3BB CRC64;

Query Match 54.3%; Score 44; DB 2; Length 314; Best Local Similarity: 47.1%; Pred. No. 6.9; Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

OY 3 EKTPLTTAAAPVYXNA 19  
|||:::|

Db 93 EKVPTSSGAPVIFVNA 109

RESULT 4

O23188 PRELIMINARY; PRT; 427 AA.

ID O23188;

DT 01-JAN-1998 (TREMblrel. 05, Created)

DT 01-MAY-1999 (TREMblrel. 10, Last sequence update)

DT 01-OCT-2000 (TREMblrel. 15, Last annotation update)

DE HYPOTHEICAL 46.8 KDA PROTEIN.

GN C7A10.390 OR ATAG36970.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; Eustoids; II; Brassicales; Brassicaceae; Arabidopsis.

OX NCBI\_TaxID=3702;

RM [1]

RP SEQUENCE FROM N.A.

RA Bevan M., Terry N., Vos P., Heljnen L., Mewes H.W., Schueller C., Chabwatiz N.;

RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RA EU Arabidopsis sequencing project;

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL: Z99707; CAB16794.1; -

DR EMBL: AL161590; CAB80363.1; -

KW Hypothetical protein.

SO SEQUENCE 427 AA; 46800 MW; 75ACC8E02D90AE54 CRC64;

Query Match 54.3%; Score 44; DB 10; Length 427; Best Local Similarity: 60.0%; Pred. No. 9.5; Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 2 EKTPLTTAAAPV 16  
|||:::|

Db 208 EEKAMTTAMQSPV 222

RESULT 5

P77924 PRELIMINARY; PRT; 506 AA.

ID P77924;

DT 01-FEB-1997 (TREMblrel. 02, Created)

DT 01-FEB-1997 (TREMblrel. 02, Last sequence update)

DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)

DE CATALASE (EC 1.11.1.6).

GN KATA.

OS Pseudomonas fluorescens.

OC Plasmid PAM10.6.

OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae; Pseudomonas.

OX NCBI\_TaxID=294;

[1]

RP SEQUENCE FROM N.A.

RC STRAIN=CB36;

RA Nurr A., Peters M.;

RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.

CC -1- FUNCTION: OCCURS IN ALMOST ALL AEROBICALLY RESPIRING ORGANISMS AND SERVES TO PROTECT CELLS FROM THE TOXIC EFFECTS OF HYDROGEN

CC PEROXIDE.  
 CC -1- CATALYTIC ACTIVITY: 2 H(2)O(2) - O(2) + 2 H(2)O.  
 CC -1- COFACTOR: HEME GROUP.  
 DR EMBL: U72068; AAB17009.1; -.  
 DR HSSP: P42321; 2CAB.  
 DR InterPro: IPR002226; Catalase.  
 DR Pfam: PF00199; catalase: 1.  
 DR PRINTS: PR00067; CATALASE.  
 DR PRODOM: PD000510; Catalase: 1.  
 DR PROSITE: PS00438; CATALASE\_2: 1.  
 KW Plasmid; Oxidoreductase; Peroxidase; Iron; Heme; Hydrogen peroxide.  
 FT ACT\_SITE 61  
 FT SEQUENCE 506 AA; 57324 MW; F1EAA728C5D41CBE CRC64;

Query Match 53.1%; Score 43; DB 2; Length 506;  
 Best Local Similarity 71.4%; Pred. No. 17;  
 Matches 10; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 5 TPLTTAAXAPVXN 18  
 DB 11 THLTTEGAPVDN 24

RESULT 6  
 ID 09DWH8 PRELIMINARY: PRT: 1240 AA.

AC 09DWH8;  
 DT 01-MAR-2001 (Tremblrel. 16, Created)  
 DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)  
 DE 01-MAR-2001 (Tremblrel. 16, Last annotation update)  
 DE PR2.  
 GN R2.  
 OS Rat cytomegalovirus (strain Maastricht).  
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
 OC Betaherpesvirinae; Muromegalovirus.  
 OX NCBI\_TaxID=79700;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-MAASTRICHT;  
 RA MEDLINE=2036325; PubMed=10906222;  
 RA Vink C., Beuken E., Bruggeman C.A.;  
 RT "Complete DNA sequence of the rat cytomegalovirus genome."  
 RT J. Virol. 74:7656-7665(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-MAASTRICHT;  
 RA MEDLINE=20473137; PubMed=11018281;  
 RA Gruljhujsen Y.K., Beuken E., Bruggeman C.A.;  
 RT "Rat cytomegalovirus R89 is a highly conserved gene which expresses a  
 RT spliced transcript."  
 RT Virus Res. 69:119-130(2000).  
 SO EMBL: AF232689; AAF99111.1; -;  
 SO SEQUENCE 1240 AA; 125612 MW; 33B6C13DC6A272B0 CRC64;

Query Match 53.1%; Score 43; DB 12; Length 1240;  
 Best Local Similarity 64.3%; Pred. No. 41;  
 Matches 9; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 2 EEKPLTTAAXAPV 15  
 DB 570 EERRELTTCASAPV 583

RESULT 7  
 ID 09RNU2 PRELIMINARY: PRT: 210 AA.

AC 09RNU2;  
 DT 01-MAY-2000 (Tremblrel. 13, Created)  
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)  
 DT 01-MAY-2000 (Tremblrel. 13, Last annotation update)  
 DE SERUM OPACITY FACTOR PRECURSOR (FRAGMENT).

GN SOF4470.  
 OS Streptococcus pyogenes.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=1314;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=4470-96; TISSUE=BLOOD;  
 RA Beall B., Gherardi G., Lovgren M., Tyrrell G., Facklam R., Fortwick B.;  
 RT "Predictions of M serotype, anti-opacity factor type, and highly  
 RT related strain sets based upon the variable emm and sof gene  
 RT sequences."  
 RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF179217; AAD55775.1; -.  
 KW Signal.  
 FT NON\_TER 1 1  
 FT SIGNAL <1 7  
 FT CHAIN 8 >210  
 FT NON\_TER 210 210  
 SO SEQUENCE 210 AA; 21220 MW; 199C89ECEF260B6A CRC64;

Query Match 51.9%; Score 42; DB 2; Length 210;  
 Best Local Similarity 57.1%; Pred. No. 10;  
 Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

OY 2 EEKPLTTAAXAPV 15  
 DB 103 EEKPTKTVTSSTPV 116

RESULT 8  
 ID 09X6X8 PRELIMINARY: PRT: 211 AA.

AC 09X6X8;  
 DT 01-NOV-1999 (Tremblrel. 12, Created)  
 DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)  
 DT 01-NOV-1999 (Tremblrel. 12, Last annotation update)  
 DE SERUM OPACITY FACTOR PRECURSOR (FRAGMENT).  
 OS Streptococcus pyogenes.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=1314;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=SS1401 TYPE PT1658;  
 RA Beall B., Gherardi G.;  
 RT "The relation of Streptococcus pyogenes sof and emm gene sequence  
 RT types to genetically distinct strain sets."  
 RT Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF154330; AAD36988.1; -.  
 KW Signal.  
 FT NON\_TER 1 1  
 FT SIGNAL <1 8  
 FT CHAIN 9 >211  
 FT NON\_TER 211 211  
 SO SEQUENCE 211 AA; 21367 MW; F5475DDC6A084FE6 CRC64;

Query Match 51.9%; Score 42; DB 2; Length 211;  
 Best Local Similarity 61.5%; Pred. No. 11;  
 Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 2 EEKPLTTAAXAP 14  
 DB 98 EEKPTKATSSAP 110

RESULT 9  
 ID 09FT45 PRELIMINARY: PRT: 436 AA.

AC 09FT45;  
 DT 01-MAR-2001 (Tremblrel. 16, Created)

DT 01-MAR-2001 (TREMBLrel. 16, last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, last annotation update)  
 DE HYPOTHETICAL 48.1 KDA PROTEIN.  
 GN T25B15.140.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eustosids II; Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Alcazar J.P., Clabault G., Cottet A., Mache R., Mexes H.W., Rudd S.,  
 RA Lemcke K., Meyer K.F.X., Quetier F., Salanoubat M.;  
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; ALI32972; CAC07928.1; -  
 DR InterPro: IPR000782; BIGH3\_fasciclin.  
 DR InterPro: IPR000903; NMT.  
 DR Pfam: PF02469; Fasciclin: 1.  
 DR PROSITE: PS00976; NMT\_2; UNKNOWN\_1.  
 DR Hypothetical protein.  
 KW SEQUENCE 436 AA; 48073 MW; 83FD94879F9855DF CRC64;  
 SO SEQUENCE

Query Match 51.9%; Score 42; DB 10; Length 436;  
 Best Local Similarity 50.0%; Pred. No. 22;  
 Matches 9; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

QY 2 EKKPLTTAAKAPVYXNA 19  
 |||||  
 DB 395 EKKPVEKKTGVYVKKKA 412

RESULT 10  
 Q9RJK9 PRELIMINARY; PRT; 487 AA.  
 AC Q9RJK9;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, last annotation update)  
 DE CATALASE (EC 1.11.1.6).  
 KATA.  
 Streptomyces coelicolor.  
 Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.  
 OX NCBI\_TaxID=1902;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-A3(2);  
 RA Murphy L., Harris D.;  
 RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-A3(2);  
 RA Thomson N.R., Parkhill J., Barrell B.G., Rajandream M.A.;  
 RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-A3(2);  
 RA Redenbach M., Kieser H.M., Denapalte D., Eichner A., Cullum J.,  
 RA Kinashi H., Hopwood D.A.;  
 RL "A set of ordered cosmid and a detailed genetic and physical map for  
 the 8 Mb Streptomyces coelicolor A3(2) chromosome.";  
 MO1. Microbiol. 21:77-96(1996).  
 CC -1- FUNCTION: OCCURS IN ALMOST ALL AEROBICALLY RESPIRING ORGANISMS AND  
 SERVES TO PROTECT CELLS FROM THE TOXIC EFFECTS OF HYDROGEN  
 PEROXIDE (BY SIMILARITY).  
 CC -1- CATALYTIC ACTIVITY: 2 H(2)O(2) = O(2) + 2 H(2)O.  
 CC EMBL; ALI21855; CAB58320.1; -

DR HSSP; P42321; 2CAE.  
 DR InterPro: IPR002226; Catalase.  
 DR Pfam: PF00199; catalase; 1.  
 DR PRINTS: PR00067; CATALASE.  
 DR PRODOM: PD000510; Catalase; 1.  
 DR PROSITE: PS00437; CATALASE\_1; 1.  
 DR PROSITE: PS00438; CATALASE\_2; 1.  
 KW Heme; Hydrogen peroxide; Iron; Oxidoreductase; Peroxidase.  
 SO SEQUENCE 487 AA; 55116 MW; 90333489EAF60B7 CRC64;

Query Match 51.9%; Score 42; DB 2; Length 487;  
 Best Local Similarity 69.2%; Pred. No. 24;  
 Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 6 PLTTAAKAPVYXN 18  
 |||||  
 DB 8 PLTTVAGAPVDPN 20

RESULT 11  
 Q9SE04 PRELIMINARY; PRT; 132 AA.  
 AC Q9SE04;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, last annotation update)  
 DE COPPER CHAPERONE HOMOLOG CCH.  
 OS Oryza sativa (Rice).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Eriaraloideae; Oryzeae; Oryza.  
 OX NCBI\_TaxID=4530;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-96369102; PubMed-9701579;  
 RA Himelblau E., Mira H., Lin S.J., Culofta V.C., Penarrubia L.,  
 RA Amasino R.M.;  
 RL "Identification of a functional homolog of the yeast copper  
 homeostasis gene ATX1 from Arabidopsis.";  
 RL Plant Physiol. 117:1227-1234(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Mira H., Penarrubia L.;  
 RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF198626; AAF15285.1; -  
 DR HSSP; P38636; ICC8.  
 DR InterPro: IPR001934; HMA.  
 DR Pfam: PF00403; HMA; 1.  
 SO SEQUENCE 132 AA; 13094 MW; 7176EF95350A8231 CRC64;

Query Match 50.6%; Score 41; DB 10; Length 132;  
 Best Local Similarity 50.0%; Pred. No. 9.9;  
 Matches 9; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

QY 2 EKKPLTTAAKAPVYXNA 19  
 |||||  
 DB 97 EAAPPTTAAEAPATAAA 114

RESULT 12  
 Q54272 PRELIMINARY; PRT; 417 AA.  
 AC Q54272;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, last annotation update)  
 DE HYPOTHETICAL ALDEHYDE-DEHYDROGENASE LIKE 43.4 KDA PROTEIN  
 (EC 1.2.1.-).  
 OS Streptomyces hygroscopicus.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;

OC Actinomycetales: Streptomycineae; Streptomycetaceae; Streptomyces.  
 OX NCBI\_TaxID=1912;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-SF1293;  
 RX MEDLINE=95309717; PubMed=779803;  
 RA Hidaka T., Hidaka M., Kuzuyama T., Seto H.;  
 RT "Sequence of a p-methyltransferase-encoding gene isolated from a  
 blalphos-producing Streptomyces hygroscopicus.";  
 RL Gene 158:149-150(1995).  
 CC -1- PATHWAY: BIALPHOS BIOSYNTHESIS.  
 CC -1- SIMILARITY: BELONGS TO THE ALDEHYDE DEHYDROGENASES FAMILY.  
 DR EMBL: D37877; BAA07116.1; -  
 DR HSSP: P05091; 1CW3.  
 DR InterPro: IPR002086; Aldehyde\_dehydr.  
 DR Pfam: PF00171; aldehyd; 1  
 DR PROSITE: PS00687; ALDEHYDE\_DEHYDR\_GLU; UNKNOWN\_1.  
 KW Hypothetical protein; Oxidoreductase.  
 FT ACT\_SITE 187 BY SIMILARITY.  
 FT ACT\_SITE 221 BY SIMILARITY.  
 SO SEQUENCE 417 AA; 43406 MW; 9E8714F654B47 CRC64;

Query Match 50.6%; Score 41; DB 2; Length 417;  
 Best Local Similarity 52.9%; Pred. No. 31;  
 Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

OY 3 EKTPLTTAAAPVXNA 19  
 DB 117 EKTPLTAFAELLTAA 133

RESULT 13  
 OY9V745 PRELIMINARY; PRT; 1012 AA.  
 AC O9V745;  
 DT 01-MAY-2000 (TReMBLrel. 13, Created)  
 DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)  
 DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)  
 DE CG18373 PROTEIN.  
 GN Drosophila melanogaster (Fruit fly).  
 OS Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Eukaryota; Metazoa; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BERKELEY;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Ceiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 Sutcliffe G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,  
 Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 Abrait J.F., Abmayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 Beeson K.Y., Benos P.V., Bertman B.P., Bhandari D., Bolshakov S.,  
 Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,  
 Burlis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 Dodson K., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 Durbin K.J., Evangelista C.C., Ferraz C., Ferrier S., Fleischmann W.,  
 Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,  
 Houston D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 Jallali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kuip D., Lai Z.,  
 Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 Liu X., Matcel B., McIntosh T.C., McLeod M.P., McPherson D.,

RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshireli A.,  
 Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclet J.M.,  
 Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 Spletter E., Spredling A.C., Stapleton M., Strong R., Sun E.,  
 Swierkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 Wang Z.-Y., Wasmann D.A., Weinstein G.M., Welzenbach J.,  
 Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of *Drosophila melanogaster*.";  
 RL Science 287:2185-2195(2000).  
 DR EMBL: AE003813; AAF58220.1; -  
 DR FLYbase: Fgn0033956; CG18373.  
 DR InterPro: IPR002052; N6\_Mtase.  
 DR PROSITE: PS00092; N6\_MTASE; UNKNOWN\_1.  
 SO SEQUENCE 1012 AA; 110009 MW; 5CF624239372C936 CRC64;

Query Match 50.6%; Score 41; DB 5; Length 1012;  
 Best Local Similarity 53.3%; Pred. No. 76;  
 Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 2 EKTPLTTAAAPV 16  
 DB 695 EEVTPYNTAPHSPIV 709

RESULT 14  
 OY9U3E3 PRELIMINARY; PRT; 171 AA.  
 AC O9U3E3;  
 DT 01-MAY-2000 (TReMBLrel. 13, Created)  
 DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)  
 DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)  
 DE F56F12.1 PROTEIN.  
 GN *Caenorhabditis elegans*.  
 OS Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;  
 OC Rhabditida; Peloderae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Ainscough R.;  
 RL Submitted (Apr-1999) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99069613; PubMed=9851916;  
 RA none;  
 RT "Genome sequence of the nematode *C. elegans*: A platform for  
 investigating biology.";  
 RL Science 282:2012-2018(1998).  
 DR EMBL: Z82273; CAB54978.1; -  
 DR InterPro: IPR003127; Sorb.  
 DR SMART: SM00459; Sorb; 1.  
 SO SEQUENCE 171 AA; 18844 MW; D782482A786F7941 CRC64;

Query Match 49.4%; Score 40; DB 5; Length 171;  
 Best Local Similarity 53.3%; Pred. No. 19;  
 Matches 8; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

OY 5 TPLTTAAAPVXNA 19  
 DB 73 TPLTTAAAPVXNA 87

RESULT 15  
 OY9U4Y3 PRELIMINARY; PRT; 212 AA.

AC 0904Y3;  
 DT 01-MAY-2000 (TRENBLREL. 13, Created)  
 DT 01-MAY-2000 (TRENBLREL. 13, Last sequence update)  
 DT 01-JUN-2001 (TRENBLREL. 17, Last annotation update)  
 DE FRUITLESS PROTEIN (FRAGMENT).  
 CN FRU.  
 OS Ceratitis capitata (Mediterranean fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 CC Tephritidae; Tephritidae; Ceratitis.  
 OX NCBI\_TaxID=7213;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Davis T., Yamamoto D.;  
 RT "Sequence of the fruitless gene in Ceratitis capitata."  
 RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.  
 EMBL: AF124047; AAF22527.1;  
 InterPro: IPR000210; BTB\_POZ.  
 DR Pfam; PF00651; BTB; 1.  
 DR PROSITE; PS50097; BTB; 1.  
 DR SMART; SM00225; BTB; 1.  
 FT NON\_TER 1 1  
 FT NON\_TER 1 1  
 SO SEQUENCE 212 AA; 22715 MW; D1B7C38FBC5AB8AB CRC64;

Query Match 49.4%; Score 40; DB 5; Length 212;  
 Best Local Similarity 52.9%; Pred. NO. 24;  
 Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

OY 3 EKTPLTTAAAPVYXNA 19  
 I::IIII I I I I  
 Db 188 ERSPLTAAMAAAVAAA 204

Search completed: February 28, 2002, 11:39:33  
 Job time: 334 sec.



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using SW model

Run On: February 28, 2002, 11:33:24 ; Search time 38.63 seconds  
(without alignments)  
36.433 Million cell updates/sec

Title: US-09-359-426C-1  
Perfect score: 48  
Sequence: 1 XEKKXLLXXXXXXVXNA 19

Scoring table: BLOSUM62  
Gap 10.0, Capex 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 08  
Maximum Match 1008  
Listing first 45 summaries

Database:

A\_Geneseq\_1101.\*  
1: /SID8/gcgdata/geneseq/geneseq/AA1980.DAT.\*  
2: /SID8/gcgdata/geneseq/geneseq/AA1981.DAT.\*  
3: /SID8/gcgdata/geneseq/geneseq/AA1982.DAT.\*  
4: /SID8/gcgdata/geneseq/geneseq/AA1983.DAT.\*  
5: /SID8/gcgdata/geneseq/geneseq/AA1984.DAT.\*  
6: /SID8/gcgdata/geneseq/geneseq/AA1985.DAT.\*  
7: /SID8/gcgdata/geneseq/geneseq/AA1986.DAT.\*  
8: /SID8/gcgdata/geneseq/geneseq/AA1987.DAT.\*  
9: /SID8/gcgdata/geneseq/geneseq/AA1988.DAT.\*  
10: /SID8/gcgdata/geneseq/geneseq/AA1989.DAT.\*  
11: /SID8/gcgdata/geneseq/geneseq/AA1990.DAT.\*  
12: /SID8/gcgdata/geneseq/geneseq/AA1991.DAT.\*  
13: /SID8/gcgdata/geneseq/geneseq/AA1992.DAT.\*  
14: /SID8/gcgdata/geneseq/geneseq/AA1993.DAT.\*  
15: /SID8/gcgdata/geneseq/geneseq/AA1994.DAT.\*  
16: /SID8/gcgdata/geneseq/geneseq/AA1995.DAT.\*  
17: /SID8/gcgdata/geneseq/geneseq/AA1996.DAT.\*  
18: /SID8/gcgdata/geneseq/geneseq/AA1997.DAT.\*  
19: /SID8/gcgdata/geneseq/geneseq/AA1998.DAT.\*  
20: /SID8/gcgdata/geneseq/geneseq/AA1999.DAT.\*  
21: /SID8/gcgdata/geneseq/geneseq/AA2000.DAT.\*  
22: /SID8/gcgdata/geneseq/geneseq/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	31	64.6	19	AAW64480	P. aeruginosa prot
2	31	64.6	19	AAW64480	Pseudomonas aerugi
3	23	47.9	892	AAAB59523	Human protein sequ
4	22	45.8	148	AAAR45152	Sequence of human
5	22	45.8	148	AAW05799	Human p154. Homo
6	22	45.8	437	AAW35264	Human adipocyte-sp
7	22	45.8	437	AAV9534	Human adipocyte-sp
8	21	43.8	130	AAAG23995	Arabidopsis thalia
9	21	43.8	134	AAAG23994	Arabidopsis thalia
10	21	43.8	151	AAAG70918	C albicans apoptos
11	21	43.8	162	AAAG23993	Arabidopsis thalia

12	21	43.8	478	22	AAAB31971	Rice glutamate 1-s
13	21	43.8	1558	21	AAAB18324	Plasmodium falcipa
14	21	43.8	1786	18	AAW24790	P. falciparum live
15	20	41.7	65	21	AAAG59124	Arabidopsis thalia
16	20	41.7	77	21	AAAG59122	Arabidopsis thalia
17	20	41.7	108	15	AAAR54307	Arabidopsis thalia
18	20	41.7	108	17	AAAM01265	Anti-HIV gp120 imm
19	20	41.7	108	21	AAV95117	VL region of HIV n
20	20	41.7	108	21	AAV98226	Anti-gp120 antibod
21	20	41.7	115	21	AAAB27999	Anti-gp120 antibod
22	20	41.7	121	22	AAAG3027	Human secreted pro
23	20	41.7	149	21	AAAG04660	C glutamic prote
24	20	41.7	159	21	AAAG04659	Arabidopsis thalia
25	20	41.7	240	22	AAAG81209	Arabidopsis thalia
26	20	41.7	257	21	AAAB18340	Mycobacterium tube
27	20	41.7	291	21	AAAG30091	Plasmodium falcipa
28	20	41.7	292	21	AAAG30090	Arabidopsis thalia
29	20	41.7	314	22	AAE06570	Arabidopsis thalia
30	20	41.7	339	18	AAW20275	Schistosoma chinens
31	20	41.7	344	22	AAAB73609	H. pylori cytoplas
32	20	41.7	350	22	AAE05988	Sporobacterium s
33	20	41.7	393	22	AAU04046	Clostridium diffic
34	20	41.7	411	20	AAV37272	Streptococcus coel
35	20	41.7	418	16	AAAR77340	Amino acid sequenc
36	20	41.7	418	16	AAAR77340	zebrafish sonic he
37	20	41.7	418	19	AAW61486	zebrafish sonic he
38	20	41.7	418	20	AAV05858	zebrafish sonic he
39	20	41.7	418	20	AAV05514	zebrafish sonic he
40	20	41.7	418	20	AAW97769	zebrafish sonic he
41	20	41.7	418	20	AAW94472	zebrafish sonic he
42	20	41.7	418	21	AAV95976	zebrafish sonic he
43	20	41.7	418	21	AAV95285	zebrafish sonic he
44	20	41.7	418	21	AAV96247	zebrafish sonic he
45	20	41.7	418	22	AAV70680	zebrafish sonic he
					AAAB60264	zebrafish sonic he

#### ALIGNMENTS

RESULT 1	AAW64480	standard; peptide: 19 AA.
ID	AAW64480	
AC	AAW64480	
XX		
XX		
DT	20-OCT-1998	(first entry)
XX		
DE	P. aeruginosa protein antigen Pa60 N-terminal peptide fragment.	
XX		
KW	Antigen: Pa60; diagnosis: detection; cystic fibrosis; vaccine;	
KW	immunogen; infection; treatment.	
XX		
OS	Pseudomonas aeruginosa.	
XX		
FH	Key	Location/Qualifiers
FT	Misc-difference 1	/label= unknown
FT	Misc-difference 12	/label= unknown
FT	Misc-difference 17	/label= unknown
FT	Misc-difference 17	/label= unknown
XX		
PN	WO9832769-A1.	
PD	30-JUL-1998.	
XX		
PF	26-JAN-1998;	98WO-GB00217.
XX		
PR	24-JAN-1997;	97GB-0001489.
XX		
PA	(AUSP-) AUSPHARM INT LTD.	
XX	(CHAP/) CHAPMAN P W.	



CC oligonucleotide comprises at least 15 nucleotides and the combination of  
CC the 5'-end sequence/3'-end sequence is selected from those defined in  
CC the specification. The primer sets can be used in antisense therapy and  
CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
CC particularly full-length cDNAs. The primers are also useful for the  
CC detection and/or diagnosis of the abnormality of the proteins encoded by  
CC the full-length cDNAs. The primers allow obtaining of the full-length  
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to  
CC AAB93893 represent human amino acid sequences; and AAH13629 to AAH13632  
CC represent oligonucleotides, all of which are used in the exemplification  
CC of the present invention.

CC Sequence 892 AA;

Query Match 47.9%; Score 23; DB 22; Length 892;  
Best Local Similarity 27.8%; Pred. No. 5.6e+02;  
Matches 5; Conservative 3; Mismatches 10; Indels 0; Gaps 0;

QY 2 EEKXXLXXXXXXXXVXNA 19  
11: : 11  
118 eerpdeeaknqlisna 135

RESULT 4

AA045152  
ID AA045152 standard; Protein: 148 AA.

AC AAR45152;

DT 20-JUN-1994 (first entry)

DE Sequence of human adipocyte polypeptide (ap) p154.

KM Adipocyte polypeptide; p154; protein 154; obesity.

OS Homo sapiens.

PN US5268295-A.

PD 07-DEC-1993.

PF 31-MAY-1991; 91US-0708038.

PR 31-MAY-1991; 91US-0708038.

PA (ALTO-) ALTON JONES CELL SCI CENT INC W.

XX Serrero G;

WI: 1993-404011/50.

DR N-PSDB: AA054135.

PT DNA encoding human or mouse adipocyte polypeptide p154 - and RNA  
PT mols. encoding p154 are used for determining susceptibility to  
PT obesity and evaluating anti-obesity drugs

PS Disclosure: columns 41-42; 49pp; English.

CC Adipose differentiation was studied using a C3H mouse teratoma-  
CC derived cell line called 1246. 1246 is bipotential, able to  
CC differentiate into adipocytes and also muscle cells. A ZAP cDNA  
CC library was constructed from fully differentiated 1246 cells. A  
CC novel adipocyte-specific polypeptide encoded by the DNA and mRNA of  
CC this cell line was discovered and termed protein 154 or p154. It is  
CC expressed in high quantities in adipogenic cell lines only after  
CC cell differentiation: it is abundant in the fat pads of normal and  
CC genetically obese mice; and its expression is 3- to 5-fold higher in  
CC the fat pads of obese animals compared to the fat pads of their  
CC littermates. A human fat cell cDNA library was hybridised with a  
CC 405 bp cDNA including the 5' end of mouse p154. The positive  
CC plaques were sequenced. Clone B4B yielded DNA of 450 bp of which

CC 474 bp contained the coding region of the human p154 DNA. Homology  
CC with the mouse p154 cDNA started at AA 50 and continued to AA 197,  
CC with 79% homology.

XX Sequence 148 AA;

Query Match 45.8%; Score 22; DB 14; Length 148;  
Best Local Similarity 27.8%; Pred. No. 1.6e+02;  
Matches 5; Conservative 3; Mismatches 10; Indels 0; Gaps 0;

QY 2 EEKXXLXXXXXXXXVXNA 19  
11: : 11  
Db 45 eerlplngpsqlvana 62

RESULT 5

AA06799  
ID AA06799 standard; Protein: 148 AA.

AC AA06799;

DT 28-JAN-1997 (first entry)

DE Human p154.

KM p154; murine; human; anti-obesity; drug evaluation; probe; antibody;  
KM susceptibility; adipocyte; efficacy.

OS Homo sapiens.

PN US5541068-A.

PD 30-JUL-1996.

PF 31-MAY-1991; 91US-0708038.

PR 31-MAY-1991; 91US-0708038.

PR 22-NOV-1993; 93US-0127995.

PA (ALTO-) ALTON JONES CELL SCI CENT INC W.

XX Serrero G;

WI: 1996-361954/36.

DR N-PSDB: AAT44456.

PT Mammalian adipocyte p154 polypeptide and related antibodies - useful  
PT in immunoassays to assess susceptibility to obesity

PS Claim 4; Column 39-40; 32pp; English.

CC The present sequence is that of human p154. The p154 mRNA is expressed  
CC in high quantities in adipogenic cell lines only after cell  
CC differentiation. The murine p154 mRNA is abundant in the fat pads of  
CC normal and genetically obese mice and its expression is 3- to 5- fold  
CC higher in the fat pads of Obese animals compared to the fat pads of  
CC their normal littermates. The DNA sequences (AAT44455-46) can be used to  
CC provide probes or for recombinant production of p154. The protein can be  
CC used for generation of antibodies. Both antibodies and probes can be used  
CC to determine the susceptibility of a subject to obesity. Also efficacy  
CC of anti-obesity drugs can be evaluated by contacting the drug to be  
CC tested with an adipocyte in vitro and measuring the amt. of p154 or mRNA  
CC produced, the efficacy of the drug being proportional to the decrease in  
CC the prodn. of the polypeptide or mRNA.

XX Sequence 148 AA;

Query Match 45.8%; Score 22; DB 17; Length 148;  
Best Local Similarity 27.8%; Pred. No. 1.6e+02;  
Matches 5; Conservative 3; Mismatches 10; Indels 0; Gaps 0;

QY 2 EEKXXLXXXXXXVXNA 19  
11: : 11  
45 eerplinqpstqlvana 62  
Db

RESULT 6  
ID AAM53264 standard; Protein: 437 AA.  
XX AAM53264;  
AC AAM53264;  
XX  
XX  
XX 01-JUL-1998 (first entry)  
XX  
XX Human adipocyte-specific differentiation-related protein.  
DE Human adipocyte-specific differentiation-related protein; HADRP;  
KW lipid metabolism; obesity; diabetes; hypercholesterolaemia;  
hyperlipidaemia; cancer.

XX Homo sapiens.  
FH Key Location/Qualifiers  
FT Misc-difference 50 /label= "unknown  
FT /note= "encoded by SCA"  
XX  
XX US5739009-A.  
XX 14-APR-1998.  
XX  
XX 12-DEC-1996; 9605-0764343.  
XX  
XX 12-DEC-1996; 9605-0764343.  
XX  
XX 12-DEC-1996; 9605-0764343.  
XX  
XX (INCY-) INCYTE PHARM INC.  
XX PA  
XX Hawkins PR, Hillman JL;  
XX  
XX WPI; 1998-260032/23.  
XX DR  
XX N-PSDB; AAV20862.  
XX  
XX Human adipocyte-specific differentiation-related protein - useful  
PT for, e.g. diagnosing or treating disorders of lipid metabolism such  
PT as obesity  
XX  
XX  
XX Claim 1; Column 31-34; 28pp; English.

CC The present sequence represents human adipocyte-specific differentiation  
CC related protein (HADRP). The DNA and protein may be used to treat or  
CC diagnose disorders of lipid metabolism, e.g. obesity, diabetes,  
CC hypercholesterolaemia or hyperlipidaemia, or to treat cancer.  
CC  
XX  
XX Sequence 437 AA;  
SQ

QY 2 EEKXXLXXXXXXVXNA 19  
11: : 11  
91 eerplinqpstqlvana 108  
Db

Query Match 45.8%; Score 22; DB 19; Length 437;  
Best Local Similarity 27.8%; Pred. NO. 5e+02;  
Matches 5; Conservative 3; Mismatches 10; Indels 0; Gaps 0;

RESULT 7  
ID AAY99534 standard; Protein: 437 AA.  
XX AAY99534;  
AC AAY99534;  
XX  
XX 27-OCT-2000 (first entry)  
XX  
XX Human adipocyte-specific differentiation-related protein ADRP.  
PE

XX  
KW Human; adipocyte-specific differentiation-related protein; ADRP;  
KW Lipid globule; chronic hepatitis; liver disease; ADRP displacement.  
XX  
XX  
XX Homo sapiens.  
OS  
XX  
XX Key Location/Qualifiers  
FH Misc-difference 50 /label= "unknown  
FT /note= "encoded by SCA"  
FT  
XX  
XX W0200031532-A1.  
XX  
XX 02-JUN-2000.  
XX  
XX 23-NOV-1999; 99MO-GB03906.  
XX  
XX 26-NOV-1998; 98GB-0025951.  
XX  
XX (MEDI-) MEDICAL RES COUNCIL.  
XX  
XX Hope G, McLauchlan J;  
XX  
XX WPI; 2000-400181/34.  
XX DR  
XX N-PSDB; AAA46136.  
XX  
XX Identifying a substance for treating or preventing a viral infection  
PT such as hepatitis C virus, comprises determining if the substance  
PT disrupts a lipid globule target sequence to lipid globule interaction  
PT  
XX  
XX Disclosure; Page 70; 74pp; English.  
XX  
XX  
XX The present sequence is the human adipocyte-specific  
CC differentiation-related protein (ADRP). It is thought that the protein  
CC is required for the maintenance of lipid droplets, but this is disrupted  
CC by infection by the hepatitis C virus, which associates with the lipid  
CC droplets and downregulates the expression of ADRP. The core protein of  
CC the hepatitis C virus, which causes chronic hepatitis and liver disease,  
CC can be used to identify substances capable of interrupting the  
CC interaction of the virus and lipid droplet, and which can be used to  
CC treat or prevent infection.  
XX  
XX  
XX Sequence 437 AA;  
SQ

QY 2 EEKXXLXXXXXXVXNA 19  
11: : 11  
91 eerplinqpstqlvana 108  
Db

Query Match 45.8%; Score 22; DB 21; Length 437;  
Best Local Similarity 27.8%; Pred. NO. 5e+02;  
Matches 5; Conservative 3; Mismatches 10; Indels 0; Gaps 0;

RESULT 8  
ID AAG23995 standard; Protein: 130 AA.  
XX AAG23995;  
AC AAG23995;  
XX  
XX 17-OCT-2000 (first entry)  
XX  
XX Arabidopsis thaliana protein fragment SEQ ID NO: 27504.  
DE  
XX Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
XX Arabidopsis thaliana.  
XX  
XX PN  
XX EP1033405-A2.  
XX

PD 06-SEP-2000.  
XX 25-FEB-2000; 2000EP-0301439.  
XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 04-MAY-1999; 99US-0132407.  
PR 05-MAY-1999; 99US-0132484.  
PR 06-MAY-1999; 99US-0132485.  
PR 07-MAY-1999; 99US-0132486.  
PR 11-MAY-1999; 99US-0132863.  
PR 14-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 18-MAY-1999; 99US-0134370.  
PR 19-MAY-1999; 99US-0134768.  
PR 20-MAY-1999; 99US-0134941.  
PR 21-MAY-1999; 99US-0135124.  
PR 24-MAY-1999; 99US-0135353.  
PR 25-MAY-1999; 99US-0135629.  
PR 27-MAY-1999; 99US-0136021.  
PR 28-MAY-1999; 99US-0136392.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 21-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147304.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 20-SEP-1999; 99US-0154739.  
PR 22-SEP-1999; 99US-0155139.

PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161922.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 43.8%; Score 21; DB 21; Length 130;  
Best Local Similarity 33.3%; Pred. No. 2.6e+02;  
Matches 6; Conservative 1; Mismatches 11; Indels 0; Gaps 0;

QY 2 EEXXXLXXXXXXVXNA 19  
|||  
98 eekatkavqclnemvtna 115

RESULT 9

ID AGG23994 standard; Protein; 134 AA.

AC AGG23994;

DT 17-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 27503.

KW Protein identification: signal transduction pathway; metabolic pathway;

KW hybridisation assay; genetic mapping; gene expression control; promoter;

KW termination sequence.

OS Arabidopsis thaliana.

PN EPI033405-A2.

PD 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439.

PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134376.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 14-JUN-1999; 99US-0138847.  
PR 16-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139889.  
PR 23-JUN-1999; 99US-0140035.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 02-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.

PR 16-JUL-1999; 99US-0144085;  
 PR 16-JUL-1999; 99US-0144086;  
 PR 19-JUL-1999; 99US-0144325;  
 PR 19-JUL-1999; 99US-0144331;  
 PR 19-JUL-1999; 99US-0144332;  
 PR 19-JUL-1999; 99US-0144333;  
 PR 19-JUL-1999; 99US-0144334;  
 PR 19-JUL-1999; 99US-0144335;  
 PR 20-JUL-1999; 99US-0144352;  
 PR 20-JUL-1999; 99US-0144632;  
 PR 20-JUL-1999; 99US-0144884;  
 PR 21-JUL-1999; 99US-0144814;  
 PR 21-JUL-1999; 99US-0145086;  
 PR 21-JUL-1999; 99US-0145088;  
 PR 22-JUL-1999; 99US-0145085;  
 PR 22-JUL-1999; 99US-0145087;  
 PR 22-JUL-1999; 99US-0145089;  
 PR 22-JUL-1999; 99US-0145192;  
 PR 23-JUL-1999; 99US-0145145;  
 PR 23-JUL-1999; 99US-0145218;  
 PR 23-JUL-1999; 99US-0145224;  
 PR 26-JUL-1999; 99US-0145276;  
 PR 27-JUL-1999; 99US-0145913;  
 PR 27-JUL-1999; 99US-0145918;  
 PR 27-JUL-1999; 99US-0145919;  
 PR 28-JUL-1999; 99US-0145951;  
 PR 02-AUG-1999; 99US-0146386;  
 PR 02-AUG-1999; 99US-0146388;  
 PR 02-AUG-1999; 99US-0146389;  
 PR 03-AUG-1999; 99US-0147038;  
 PR 04-AUG-1999; 99US-0147204;  
 PR 04-AUG-1999; 99US-0147302;  
 PR 05-AUG-1999; 99US-0147192;  
 PR 05-AUG-1999; 99US-0147260;  
 PR 06-AUG-1999; 99US-0147303;  
 PR 06-AUG-1999; 99US-0147416;  
 PR 09-AUG-1999; 99US-0147493;  
 PR 09-AUG-1999; 99US-0147935;  
 PR 10-AUG-1999; 99US-0148171;  
 PR 11-AUG-1999; 99US-0148319;  
 PR 12-AUG-1999; 99US-0148341;  
 PR 13-AUG-1999; 99US-0148565;  
 PR 13-AUG-1999; 99US-0148684;  
 PR 16-AUG-1999; 99US-0149368;  
 PR 17-AUG-1999; 99US-0149175;  
 PR 18-AUG-1999; 99US-0149426;  
 PR 20-AUG-1999; 99US-0149722;  
 PR 20-AUG-1999; 99US-0149723;  
 PR 20-AUG-1999; 99US-0149929;  
 PR 23-AUG-1999; 99US-0149902;  
 PR 23-AUG-1999; 99US-0149930;  
 PR 25-AUG-1999; 99US-0150566;  
 PR 26-AUG-1999; 99US-0150884;  
 PR 27-AUG-1999; 99US-0151065;  
 PR 27-AUG-1999; 99US-0151066;  
 PR 27-AUG-1999; 99US-0151080;  
 PR 30-AUG-1999; 99US-0151303;  
 PR 31-AUG-1999; 99US-0151338;  
 PR 01-SEP-1999; 99US-0151930;  
 PR 07-SEP-1999; 99US-0152363;  
 PR 10-SEP-1999; 99US-0153070;  
 PR 13-SEP-1999; 99US-0153758;  
 PR 15-SEP-1999; 99US-0154018;  
 PR 16-SEP-1999; 99US-0154039;  
 PR 20-SEP-1999; 99US-0154779;  
 PR 22-SEP-1999; 99US-0155139;  
 PR 23-SEP-1999; 99US-0155486;  
 PR 24-SEP-1999; 99US-0155659;  
 PR 28-SEP-1999; 99US-0156458;  
 PR 29-SEP-1999; 99US-0156596;  
 PR 04-OCT-1999; 99US-0157117;  
 PR 05-OCT-1999; 99US-0157753;  
 PR 06-OCT-1999; 99US-0157865;

PR 07-OCT-1999; 99US-0158029;  
 PR 08-OCT-1999; 99US-0158232;  
 PR 12-OCT-1999; 99US-0158369;  
 PR 13-OCT-1999; 99US-0158293;  
 PR 13-OCT-1999; 99US-0159294;  
 PR 13-OCT-1999; 99US-0159295;  
 PR 14-OCT-1999; 99US-0159329;  
 PR 14-OCT-1999; 99US-0159330;  
 PR 14-OCT-1999; 99US-0159331;  
 PR 14-OCT-1999; 99US-0159637;  
 PR 14-OCT-1999; 99US-0159638;  
 PR 18-OCT-1999; 99US-0159584;  
 PR 21-OCT-1999; 99US-0160741;  
 PR 21-OCT-1999; 99US-0160767;  
 PR 21-OCT-1999; 99US-0160768;  
 PR 21-OCT-1999; 99US-0160770;  
 PR 21-OCT-1999; 99US-0160814;  
 PR 21-OCT-1999; 99US-0160815;  
 PR 22-OCT-1999; 99US-0160980;  
 PR 22-OCT-1999; 99US-0160981;  
 PR 22-OCT-1999; 99US-0160989;  
 PR 25-OCT-1999; 99US-0161404;  
 PR 25-OCT-1999; 99US-0161405;  
 PR 25-OCT-1999; 99US-0161406;  
 PR 26-OCT-1999; 99US-0161359;  
 PR 26-OCT-1999; 99US-0161360;  
 PR 26-OCT-1999; 99US-0161361;  
 PR 28-OCT-1999; 99US-0161920;  
 PR 28-OCT-1999; 99US-0161992;  
 PR 28-OCT-1999; 99US-0161993;  
 PR 29-OCT-1999; 99US-0162142;

Query Match 43.8%; Score 21; DB 21; Length 134;  
 Best Local Similarity 33.3%; Pred. No. 2.7e+02;  
 Matches 6; Conservative 1; Mismatches 11; Indels 0; Gaps 0;

Oy 2 EEKXXLXXXXXXVXNA 19  
 ||| :|||  
 Db 102 eekatkavqclnewvtna 119

RESULT 10  
 AAG70918  
 ID AAG70918 standard; Protein; 151 AA.

XX AAG70918;  
 AC  
 XX  
 DT 27-JUL-2001 (first entry)  
 XX  
 DE C albicans apoptosis associated protein #90.

XX  
 KM Yeast; fungus; apoptosis; infection; proliferative disease;  
 KW vaccine; autoimmune disease; Ischemia; neurodegeneration.

XX  
 OS Candida albicans.

XX  
 PN WO200102550-A2.

XX  
 PD 11-JAN-2001.

XX  
 PF 03-JUL-2000; 2000WO-BE00077.

XX  
 PR 01-JUL-1999; 99EP-0870141.

XX  
 PA (JANC) JANSSEN PHARM NV.

XX  
 PI Contreras RH, De Backer MD, Luyten WHML, Malcorps IKL;  
 PI Nelsissen BJM, Reekmans RJ;

XX  
 DR WPT: 2001-367042/38.

DR  
 N-PSDB: AAH29954.

PT Yeast and fungal nucleic acids encoding proteins involved in a pathway

PT leading to programmed cell death, useful for treating proliferative  
PT disorders, yeast and fungal infections, or for preventing apoptosis in  
PT certain diseases  
PS Claim 24; Fig 2; 218bp; English.  
XX  
CC The present invention provides the protein and coding sequences of a  
CC number of apoptosis associated proteins from the yeast *Saccharomyces*  
CC *cerevisiae* and the fungus *Candida albicans*. These can be used to identify  
CC treatments for fungal and yeast infections, for proliferative diseases  
CC and for apoptosis related diseases such as autoimmune diseases, ischemia  
CC and neurodegeneration. The present sequence is one of the *C. albicans*  
CC proteins of the invention.  
XX  
XX Sequence 151 AA;  
Query Match 43.8%; Score 21; DB 22; Length 151;  
Best Local Similarity 31.2%; Pred. No. 3.1e+02;  
Matches 5; Conservative 1; Mismatches 10; Indels 0; Gaps 0;  
CY 3 EKXLLXXXXXXVXN 18  
Db 116 ekavignidasalvin 131  
RESULT 11  
AAG23993  
ID AAG23993 standard; Protein: 162 AA.  
XX  
XX AAG23993;  
XX  
DT 17-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 27502.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
XX EP1033405-A2.  
XX  
PD 06-SEP-2000.  
25-FEB-2000; 2000EP-0301439.  
25-FEB-1999; 9905-0121825.  
PR 05-MAR-1999; 9905-0123180.  
PR 09-MAR-1999; 9905-0123548.  
PR 23-MAR-1999; 9905-0125788.  
PR 25-MAR-1999; 9905-0126264.  
PR 29-MAR-1999; 9905-0126785.  
PR 01-APR-1999; 9905-0127462.  
PR 06-APR-1999; 9905-0128234.  
PR 08-APR-1999; 9905-0128714.  
PR 16-APR-1999; 9905-0129845.  
PR 19-APR-1999; 9905-0130077.  
PR 21-APR-1999; 9905-0130449.  
PR 23-APR-1999; 9905-0130510.  
PR 23-APR-1999; 9905-0130891.  
PR 28-APR-1999; 9905-0131449.  
PR 30-APR-1999; 9905-0132048.  
PR 30-APR-1999; 9905-0132407.  
PR 04-MAY-1999; 9905-0132484.  
PR 05-MAY-1999; 9905-0132485.  
PR 06-MAY-1999; 9905-0132486.  
PR 07-MAY-1999; 9905-0132487.  
PR 11-MAY-1999; 9905-0134256.  
PR 14-MAY-1999; 9905-0134218.  
PR 14-MAY-1999; 9905-0134219.

PR 14-MAY-1999; 9905-0134221.  
PR 14-MAY-1999; 9905-0134370.  
PR 18-MAY-1999; 9905-0134768.  
PR 19-MAY-1999; 9905-0134941.  
PR 20-MAY-1999; 9905-0135124.  
PR 21-MAY-1999; 9905-0135353.  
PR 24-MAY-1999; 9905-0135629.  
PR 25-MAY-1999; 9905-0136021.  
PR 27-MAY-1999; 9905-0136392.  
PR 28-MAY-1999; 9905-0136782.  
PR 01-JUN-1999; 9905-0137222.  
PR 03-JUN-1999; 9905-0137528.  
PR 04-JUN-1999; 9905-0137502.  
PR 07-JUN-1999; 9905-0137724.  
PR 08-JUN-1999; 9905-0138094.  
PR 10-JUN-1999; 9905-0138540.  
PR 10-JUN-1999; 9905-0138847.  
PR 14-JUN-1999; 9905-0139119.  
PR 16-JUN-1999; 9905-0139452.  
PR 16-JUN-1999; 9905-0139453.  
PR 17-JUN-1999; 9905-0139492.  
PR 18-JUN-1999; 9905-0139454.  
PR 18-JUN-1999; 9905-0139455.  
PR 18-JUN-1999; 9905-0139456.  
PR 18-JUN-1999; 9905-0139457.  
PR 18-JUN-1999; 9905-0139458.  
PR 18-JUN-1999; 9905-0139459.  
PR 18-JUN-1999; 9905-0139460.  
PR 18-JUN-1999; 9905-0139461.  
PR 18-JUN-1999; 9905-0139462.  
PR 18-JUN-1999; 9905-0139463.  
PR 18-JUN-1999; 9905-0139750.  
PR 18-JUN-1999; 9905-0139753.  
PR 21-JUN-1999; 9905-0139817.  
PR 22-JUN-1999; 9905-0139859.  
PR 23-JUN-1999; 9905-0140353.  
PR 23-JUN-1999; 9905-0140354.  
PR 24-JUN-1999; 9905-0140695.  
PR 28-JUN-1999; 9905-0140823.  
PR 29-JUN-1999; 9905-0140991.  
PR 30-JUN-1999; 9905-0141287.  
PR 01-JUL-1999; 9905-0141842.  
PR 01-JUL-1999; 9905-0142154.  
PR 02-JUL-1999; 9905-0142055.  
PR 06-JUL-1999; 9905-0142390.  
PR 08-JUL-1999; 9905-0142803.  
PR 09-JUL-1999; 9905-0142920.  
PR 12-JUL-1999; 9905-0142977.  
PR 13-JUL-1999; 9905-0143542.  
PR 14-JUL-1999; 9905-0143624.  
PR 15-JUL-1999; 9905-0144005.  
PR 16-JUL-1999; 9905-0144085.  
PR 16-JUL-1999; 9905-0144086.  
PR 16-JUL-1999; 9905-0144325.  
PR 19-JUL-1999; 9905-0144331.  
PR 19-JUL-1999; 9905-0144332.  
PR 19-JUL-1999; 9905-0144333.  
PR 19-JUL-1999; 9905-0144334.  
PR 19-JUL-1999; 9905-0144335.  
PR 20-JUL-1999; 9905-0144352.  
PR 20-JUL-1999; 9905-0144632.  
PR 20-JUL-1999; 9905-0144884.  
PR 21-JUL-1999; 9905-0144814.  
PR 21-JUL-1999; 9905-0145086.  
PR 21-JUL-1999; 9905-0145088.  
PR 22-JUL-1999; 9905-0145085.  
PR 22-JUL-1999; 9905-0145087.  
PR 22-JUL-1999; 9905-0145089.  
PR 22-JUL-1999; 9905-0145192.  
PR 23-JUL-1999; 9905-0145145.  
PR 23-JUL-1999; 9905-0145218.  
PR 23-JUL-1999; 9905-0145224.  
PR 26-JUL-1999; 9905-0145276.



```

PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 43.8%; Score 21; DB 21; length 162.
Best Local Similarity 33.3%; Pred. No. 3,36+02.
Matches 6; Conservative 1; Mismatches 11; Indels 0; Gaps 0;

Oy 2 EEKXXLXXXXXXVXNA 19
    |||
Db 130 eekakavqclnemvtna 147

RESULT 12
AAB31971
ID AAB31971 standard; Protein: 478 AA.
XX
AC AAB31971:
XX
DT 15-MAY-2001 (first entry)
XX
DE Rice glutamate 1-semialdehyde (GSA) aminotransferase.
XX
KW Glu-tRNA reductase; aminolevulinic acid; 5-aminolevulinic acid; ALA;
KW tetrapyrrolic pigment; chlorophyll; transgenic plant; GSA;
KW 1-semialdehyde aminotransferase.
XX
OS Oryza sativa.
XX
FH Key Location/Qualifiers
FT Misc-difference 319 /note= "Ile encoded by TC"
FT FT Misc-difference 981 /note= "unknown residue encoded by C"
XX
PN W0200109304-A2.
PD 08-FEB-2001.
XX
PF 28-JUL-2000; 2000WO-US21008.
XX
PR 30-JUL-1999; 99US-0146600.
XX
PA (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
PA Cahoon RE, Gutteridge S, Harvell LT, Rafalski JA, Tao Y, Weng Z;
XX
DR WP1: 2001-159865/16.
XX
DR N-PSDB: AAF54825.

New nucleic acids encoding an aminolevulinic acid biosynthetic enzyme
useful for producing large amounts of the encoded polypeptides used in
screening compounds for potential herbicidal activity -

Disclosure: Page 73-75; 77pp; English.

The present sequence represents a glutamate 1-semialdehyde (GSA)
aminotransferase which is an aminolevulinic acid biosynthetic enzyme.
Glu-tRNA reductase converts Glu-tRNA-Glu to GSA with the concomitant
release of tRNA-Glu. GSA aminotransferase then converts GSA to
5-aminolevulinic acid (ALA). ALA is used in the biosynthesis of
tetrapyrrolic pigments such as chlorophyll. Nucleic acids encoding
aminolevulinic acid biosynthetic enzymes are useful for isolating cDNAs
and genes encoding homologous proteins from the same or other plant
species, for creating transgenic plants in which the polypeptides are
present at higher or lower levels than normal or in cell types or
developmental stages in which they are not normally found, for

```

```

PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 43.8%; Score 21; DB 21; length 162.
Best Local Similarity 33.3%; Pred. No. 3,36+02.
Matches 6; Conservative 1; Mismatches 11; Indels 0; Gaps 0;

Oy 2 EEKXXLXXXXXXVXNA 19
    |||
Db 130 eekakavqclnemvtna 147

RESULT 12
AAB31971
ID AAB31971 standard; Protein: 478 AA.
XX
AC AAB31971:
XX
DT 15-MAY-2001 (first entry)
XX
DE Rice glutamate 1-semialdehyde (GSA) aminotransferase.
XX
KW Glu-tRNA reductase; aminolevulinic acid; 5-aminolevulinic acid; ALA;
KW tetrapyrrolic pigment; chlorophyll; transgenic plant; GSA;
KW 1-semialdehyde aminotransferase.
XX
OS Oryza sativa.
XX
FH Key Location/Qualifiers
FT Misc-difference 319 /note= "Ile encoded by TC"
FT FT Misc-difference 981 /note= "unknown residue encoded by C"
XX
PN WC200109304-A2.
PD 08-FEB-2001.
XX
PF 28-JUL-2000; 2000WO-US21008.
XX
PR 30-JUL-1999; 99US-0146600.
XX
PA (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
PA Cahoon RE, Gutteridge S, Harvell LT, Rafalski JA, Tao Y, Weng Z;
DR WP1: 2001-159865/16.
DR N-PSDB: AAF54825.
XX

New nucleic acids encoding an aminolevulinic acid biosynthetic enzyme
useful for producing large amounts of the encoded polypeptides used in
screening compounds for potential herbicidal activity -

Disclosure: Page 73-75; 77pp; English.
XX
XX The present sequence represents a glutamate 1-semialdehyde (GSA)
XX aminotransferase which is an aminolevulinic acid biosynthetic enzyme.
XX Glu-tRNA reductase converts Glu-tRNA-Glu to GSA with the concomitant
XX release of tRNA-Glu. GSA aminotransferase then converts GSA to
XX 5-aminolevulinic acid (ALA). ALA is used in the biosynthesis of
XX tetrapyrrolic pigments such as chlorophyll. Nucleic acids encoding
XX aminolevulinic acid biosynthetic enzymes are useful for isolating cDNAs
XX and genes encoding homologous proteins from the same or other plant
XX species, for creating transgenic plants in which the polypeptides are
XX present at higher or lower levels than normal or in cell types or
XX developmental stages in which they are not normally found, for

```

CC overexpression in bacterial or yeast hosts to efficiently produce  
CC large amounts of the encoded polypeptides which could then be used for  
CC screening different compounds for potential herbicidal activity, and  
CC as hybridisation probes and amplification primers.

XX Sequence 478 AA;

Query Match 43.8%; Score 21; DB 22; Length 478;  
Best Local Similarity 33.3%; Pred. No. 1e+03;  
Matches 6; Conservative 1; Mismatches 11; Indels 0; Gaps 0;

QY 2 EKKXXLXXXXXXVYVNA 19  
|||  
246 enkqgiaavflepyvyna 263

RESULT 13  
ID AAB18324 standard; Protein; 1558 AA.

XX AAB18324;

XX 07-NOV-2000 (first entry)

DE Plasmodium falciparum chromosome 2 related protein SEQ ID NO:182.

KW Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;  
KM antimalarial; malaria; protozoacide; infection; insecticide.

XX Plasmodium falciparum.

PN MO200025728-A2.

PD 11-MAY-2000.

PF 05-NOV-1999; 99WO-US26796.

PR 05-NOV-1998; 98US-0107131.

XX (HOFF/) HOFFMAN S.  
PA (CARU/) CARUCCI D.  
PA (GARD/) GARDNER M.  
PA (VENT/) VENTER J C.

Hoffman S, Carucci D, Gardner M, Venter JC;

WPI; 2000-365347/31.

XX Proteins encoded by chromosome 2 of the human malarial parasite,  
PT Plasmodium falciparum, useful as antimalarial vaccines and in the  
PT diagnosis of P. falciparum infection -

PS Disclosure; Page 410-414; 577pp; English.

XX The present invention describes proteins and their fragments (I) encoded  
CC by chromosome 2 of the human malarial parasite, Plasmodium falciparum.  
CC Also described are: (I) nucleotide sequences (II) encoding (I); and (2) '  
CC vaccines against P. falciparum infection comprising (I) or (II).  
CC (I) and (II) are useful for the development of vaccines against  
CC P. falciparum infection. (I) and polyclonal antisera or a monoclonal  
CC antibody raised to immunogens comprising the sequences of (I), are  
CC useful in the detection of infection with P. falciparum. Furthermore,  
CC (I) (especially when they are rins or secreted or membrane proteins)  
CC can aid the identification of drugs to treat or prevent P. falciparum  
CC infection, or they can be used to identify drug resistance in  
CC P. falciparum. Sequencing of the Plasmodium chromosome 2 and the  
CC subsequent identification of proteins encoded by it will help to expand  
CC our understanding of parasite biology, a process hampered by the  
CC complexity of the parasitic lifecycle, and provide new targets for  
CC vaccine and drug development. Parasite resistance to drugs and mosquito  
CC resistance to insecticides have led to a resurgence of malaria in many  
CC parts of the world, and there is a pressing need for vaccines and new

CC drugs. AAW70078 to AAW70287 and AAB18144 to AAB18352 represent nucleotide  
CC and protein sequences given in the present invention, but which are not  
CC specifically mentioned within the specification.

XX Sequence 1558 AA;

Query Match 43.8%; Score 21; DB 21; Length 1558;  
Best Local Similarity 29.4%; Pred. No. 3.6e+03;  
Matches 5; Conservative 2; Mismatches 10; Indels 0; Gaps 0;

QY 2 EKKXXLXXXXXXVYVXN 18  
|||  
806 eekvdlnenvvssldn 822

RESULT 14  
ID AAW24790 standard; Protein; 1786 AA.

XX AAW24790;

XX 08-OCT-1997 (first entry)

DE P. falciparum liver stage antigen-3.

KW Plasmodium falciparum; pre-erythrocyte; liver stage antigen; serum;  
KM prophylaxis; Thai strain; gene organisation; exon; intron; hydrophobic;  
KM glycosyl-phosphatidylinositol membrane anchoring sequence; antibody;

XX vaccine; immunotherapy; malaria.

OS Plasmodium falciparum.

XX Key Location/Qualifiers

FT Region 223..278 /note= "repeat region 1"

FT Region 279..818 /note= "repeat region 2"

FT Region 1537..1576 /note= "repeat region 3"

XX MO9641877-A2.

XX 27-DEC-1996.

PF 12-JUN-1996; 96WO-FR00894.

PR 13-JUN-1995; 95FR-0007007.

XX (INSP ) INST PASTEUR.

XX Daubersies P, Drullhe P;

XX WPI; 1997-065464/06.

XX N-PSDB; AAT78868.

PT Plasmodium falciparum poly(peptide)s and related nucleic acids -  
PT derived from the liver stage antigen-3, useful for malaria vaccine  
PT prodn. and diagnosis

PS Claim 1; Fig 2A-I; 69pp; French.

XX This sequence corresponds to a Plasmodium falciparum strain K1  
CC pre-erythrocytic liver stage antigen-3 (LSA-3) protein. The encoding  
CC gene sequence was isolated by screening a P. falciparum strain T9/96  
CC library with serum from a missionary treated by prophylaxis (for strain  
CC T6/96 see FR9101286). Of 20 clones isolated, clone 7295 was used to  
CC screen a library generated from Thai strain K1. One clone contained a  
CC 6.85 kb insert including the genomic sequence AAT78867. The gene  
CC comprises a 1.8 kb region encoding 3 major blocks of tetrapeptide  
CC repeats (especially the amino acid sequence VEEV, VEEV, VAPV, VAPV,  
CC etc) and a 3' hydrophobic region corresponding to a  
CC glycosyl-phosphatidyl- inositol membrane anchoring sequence. The

CC Invention relates to new polypeptides of at least 10 amino acids derived  
CC from the USA-3 protein with the exception of the peptides AAM24791-4.  
CC The USA-3 peptides can be used to raise antibodies and as vaccines for  
CC immunotherapy of malaria.  
XX  
SQ Sequence 1786 AA:

Query Match 43.8%; Score 21; DB 18; Length 1786;  
Best Local Similarity 29.4%; Pred. No. 4.2e+03;  
Matches 5; Conservative 2; Mismatches 10; Indels 0; Gaps 0;

Oy 2 EKKXXLXXXXXXVXN 18  
Db 1038 eekvdlnenvsldh 1054

RESULT 15  
AAC59124  
ID AAC59124 standard; Protein: 65 AA.

XX AAC59124:

XX 18-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 76445.

XX Protein Identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.

XX Arabidopsis thaliana.

XX EPI033405-A2.

XX PD 06-SEP-2000.

XX PF 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 11-MAY-1999; 99US-0132863.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.

PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144086.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145213.  
PR 27-JUL-1999; 99US-0145218.  
PR 27-JUL-1999; 99US-0145219.  
PR 28-JUL-1999; 99US-0145551.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.

PR 04-AUG-1999; 99US-0147302.  
 PR 05-AUG-1999; 99US-0147192.  
 PR 05-AUG-1999; 99US-0147260.  
 PR 06-AUG-1999; 99US-0147303.  
 PR 06-AUG-1999; 99US-0147416.  
 PR 09-AUG-1999; 99US-0147493.  
 PR 09-AUG-1999; 99US-0147935.  
 PR 10-AUG-1999; 99US-0148171.  
 PR 11-AUG-1999; 99US-0148319.  
 PR 12-AUG-1999; 99US-0148341.  
 PR 13-AUG-1999; 99US-0148565.  
 PR 13-AUG-1999; 99US-0148684.  
 PR 16-AUG-1999; 99US-0149368.  
 PR 17-AUG-1999; 99US-0149175.  
 PR 18-AUG-1999; 99US-0149426.  
 PR 20-AUG-1999; 99US-0149722.  
 PR 20-AUG-1999; 99US-0149723.  
 PR 20-AUG-1999; 99US-0149928.  
 PR 23-AUG-1999; 99US-0149902.  
 PR 23-AUG-1999; 99US-0149930.  
 PR 25-AUG-1999; 99US-0150566.  
 PR 26-AUG-1999; 99US-0150884.  
 PR 27-AUG-1999; 99US-0151065.  
 PR 27-AUG-1999; 99US-0151066.  
 PR 27-AUG-1999; 99US-0151080.  
 PR 30-AUG-1999; 99US-0151303.  
 PR 31-AUG-1999; 99US-0151438.  
 PR 01-SEP-1999; 99US-0151930.  
 PR 07-SEP-1999; 99US-0152363.  
 PR 10-SEP-1999; 99US-0153070.  
 PR 13-SEP-1999; 99US-0153758.  
 PR 15-SEP-1999; 99US-0154018.  
 PR 16-SEP-1999; 99US-0154039.  
 PR 20-SEP-1999; 99US-0154779.  
 PR 22-SEP-1999; 99US-0155139.  
 PR 23-SEP-1999; 99US-0155486.  
 PR 24-SEP-1999; 99US-0155659.  
 PR 28-SEP-1999; 99US-0156458.  
 PR 29-SEP-1999; 99US-0156596.  
 PR 04-OCT-1999; 99US-0157117.  
 PR 05-OCT-1999; 99US-0157753.  
 PR 06-OCT-1999; 99US-0157865.  
 PR 07-OCT-1999; 99US-0158028.  
 PR 08-OCT-1999; 99US-0158232.  
 PR 12-OCT-1999; 99US-0158369.  
 PR 13-OCT-1999; 99US-0159293.  
 PR 13-OCT-1999; 99US-0159294.  
 PR 13-OCT-1999; 99US-0159295.  
 PR 14-OCT-1999; 99US-0159328.  
 PR 14-OCT-1999; 99US-0159330.  
 PR 14-OCT-1999; 99US-0159331.  
 PR 14-OCT-1999; 99US-0159637.  
 PR 14-OCT-1999; 99US-0159638.  
 PR 18-OCT-1999; 99US-0159584.  
 PR 21-OCT-1999; 99US-0160741.  
 PR 21-OCT-1999; 99US-0160767.  
 PR 21-OCT-1999; 99US-0160770.  
 PR 21-OCT-1999; 99US-0160776.  
 PR 21-OCT-1999; 99US-0160814.  
 PR 21-OCT-1999; 99US-0160815.  
 PR 22-OCT-1999; 99US-0160980.  
 PR 22-OCT-1999; 99US-0160981.  
 PR 22-OCT-1999; 99US-0160989.  
 PR 22-OCT-1999; 99US-0161404.  
 PR 25-OCT-1999; 99US-0161405.  
 PR 25-OCT-1999; 99US-0161406.  
 PR 26-OCT-1999; 99US-0161359.  
 PR 26-OCT-1999; 99US-0161360.  
 PR 26-OCT-1999; 99US-0161361.  
 PR 28-OCT-1999; 99US-0161920.  
 PR 28-OCT-1999; 99US-0161992.  
 PR 28-OCT-1999; 99US-0161993.  
 PR 29-OCT-1999; 99US-0162142.

Query Match 41.7%; Score 20; DB 21; Length 65;  
 Best Local Similarity 33.3%; Pred. No. 2.4e+02;  
 Matches 5; Conservative 1; Mismatches 9; Indels 0; Gaps 0;  
 Oy 2 EEXXXLXXXXXXV 16  
 ||: |  
 Db 35 eertslavavagvv 49

Search completed: February 28, 2002, 11:34:38  
 Job time: 74 sec



**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 28, 2002, 11:33:24 : Search time 19.88 seconds  
(without alignments)  
21.507 Million cell updates/sec

Title: US-09-359-426c-1

Perfect score: 48  
Sequence: 1 XEKKXXLXXXXXXVXNA 19

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/2/1aa/5a\_COMB.pep:\*  
2: /cgn2\_6/ptodata/2/1aa/5b\_COMB.pep:\*  
3: /cgn2\_6/ptodata/2/1aa/6a\_COMB.pep:\*  
4: /cgn2\_6/ptodata/2/1aa/6b\_COMB.pep:\*  
5: /cgn2\_6/ptodata/2/1aa/pctus\_COMB.pep:\*  
6: /cgn2\_6/ptodata/2/1aa/backfilest1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	22	45.8	148	1	US-07-708-038-4
2	22	45.8	148	1	US-08-127-995-4
3	22	45.8	437	1	US-08-764-343-1
4	22	45.8	437	2	US-08-989-925-3
5	21	43.8	630	4	US-08-973-462-9
6	21	43.8	1786	4	US-08-973-462-8
7	20	41.7	108	1	US-08-899-575-86
8	20	41.7	108	1	US-08-899-575-86
9	20	41.7	108	1	US-08-899-575-86
10	20	41.7	108	5	PCT-US95-08743-66
11	20	41.7	148	3	US-08-946-329A-63
12	20	41.7	314	4	US-09-475-316A-118
13	20	41.7	418	1	US-08-176-427B-10
14	20	41.7	418	1	US-08-356-606A-12
15	20	41.7	418	3	US-08-946-329A-18
16	20	41.7	418	4	US-08-567-357A-18
17	20	41.7	418	4	US-08-460-900C-12
18	20	41.7	418	4	US-08-729-743A-18
19	20	41.7	418	4	US-08-757-230A-7
20	20	41.7	418	4	US-08-674-509B-12
21	20	41.7	418	4	US-08-954-698-12
22	20	41.7	418	4	US-08-349-498-18
23	20	41.7	418	5	PCT-US95-15463-18
24	20	41.7	418	5	PCT-US95-15923-18
25	19	39.6	35	4	US-09-248-588-26
26	19	39.6	101	1	US-08-241-853-16
27	19	39.6	101	2	US-08-850-917-16

## ALIGNMENTS

28	19	39.6	226	1	US-07-929-198-4	Sequence 4, Appl1
29	19	39.6	288	3	US-08-312-949-4	Sequence 4, Appl1
30	19	39.6	288	3	US-08-446-201-4	Sequence 4, Appl1
31	19	39.6	289	1	US-08-072-070-4	Sequence 4, Appl1
32	19	39.6	289	1	US-08-469-434-4	Sequence 4, Appl1
33	19	39.6	289	1	US-08-214-222-4	Sequence 4, Appl1
34	19	39.6	289	2	US-08-467-852A-5	Sequence 4, Appl1
35	19	39.6	289	2	US-08-468-718-4	Sequence 5, Appl1
36	19	39.6	289	2	US-08-247-491A-5	Sequence 5, Appl1
37	19	39.6	299	3	US-08-923-856-1	Sequence 5, Appl1
38	19	39.6	299	3	US-09-216-294-1	Sequence 1, Appl1
39	19	39.6	331	4	US-09-457-046B-59	Sequence 59, Appl1
40	19	39.6	498	1	US-08-470-202-60	Sequence 60, Appl1
41	19	39.6	498	1	US-08-471-770-60	Sequence 60, Appl1
42	19	39.6	498	2	US-08-468-059-60	Sequence 60, Appl1
43	19	39.6	498	2	US-08-511-485-13	Sequence 13, Appl1
44	19	39.6	498	4	US-09-109-916-60	Sequence 60, Appl1
45	19	39.6	615	4	US-08-676-444-44	Sequence 44, Appl1

RESULT 1  
US-07-708-038-4  
Sequence 4, Application US/07708038  
Patent No. 5268295  
GENERAL INFORMATION:  
APPLICANT: Serrero, Glanette  
TITLE OF INVENTION: MAMMALIAN ADIPOCYTE PROTEIN, NUCLEIC  
TITLE OR INVENTION: ACIDS CODING  
THEREFOR AND USES THEREOF  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Browdy and Neimark  
STREET: 419 Seventh Street, NW  
CITY: Washington  
STATE: DC  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07708,038  
FILING DATE: 19910531  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Livnat, Shmuel  
REGISTRATION NUMBER: 33,949  
REFERENCE/DOCKET NUMBER: SERRERO-2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-628-5197  
TELEFAX: 202-737-3528  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 148 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-07-708-038-4  
Query Match 45.88; Score 22; DB 1; Length 148;  
Best Local Similarity 27.88; Pred. No. 66;  
Matches 5; Conservative 3; Mismatches 10; Indels 0; Gaps 0;

RESULT 2  
US-08-127-995-4  
Sequence 4, Application US/08127995  
Patent No. 5541068  
GENERAL INFORMATION:  
APPLICANT: Sertero, Ginette  
TITLE OF INVENTION: MAMMALIAN ADIPOCYTE PROTEIN, NUCLEIC  
TITLE OF INVENTION: ACIDS CODING THEREFOR AND USES THEREOF  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Browdy and Neimark  
STREET: 419 Seventh Street, NW  
CITY: Washington  
STATE: DC  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentln Release #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/127,995  
FILING DATE: 15-NOV-1993  
CLASSIFICATION: 1.435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/708,038  
FILING DATE: 31-MAY-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Townsend, G. Kevin  
REGISTRATION NUMBER: 34,033  
REFERENCE/DOCKET NUMBER: SERTERO-2A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-628-5197  
TELEFAX: 202-737-3528  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 148 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-127-995-4

Query Match 45.8%; Score 22; DB 1; Length 148;  
Best Local Similarity 27.8%; Pred. No. 66;  
Matches 5; Conservative 3; Mismatches 10; Indels 0; Gaps 0;

2 EERXXLXXXXXXVXNA 19  
11: : 111  
DB 45 EERLPINQSTQIVANA 62

RESULT 3  
US-08-764-343-1  
Sequence 1, Application US/0876433  
Patent No. 5739009  
GENERAL INFORMATION:  
APPLICANT: Hillman, Jennifer L.  
TITLE OF INVENTION: NOVEL ADIPOCYTE-SPECIFIC  
TITLE OF INVENTION: DIFFERENTIATION-RELATED PROTEIN  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: INCYTE PHARMACEUTICALS, INC.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: US  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
OPERATING SYSTEM: DOS

SOFTWARE: FastSP Version 1.5  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/764,343  
FILING DATE: Filed Herewith  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0167 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 437 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
IMMEDIATE SOURCE:  
CLONE: Consensus  
US-08-764-343-1

Query Match 45.8%; Score 22; DB 1; Length 437;  
Best Local Similarity 27.8%; Pred. No. 2e+02;  
Matches 5; Conservative 3; Mismatches 10; Indels 0; Gaps 0;

QY 2 EERXXLXXXXXXVXNA 19  
11: : 111  
DB 91 EERLPINQSTQIVANA 108

RESULT 4  
US-08-989-925-3  
Sequence 3, Application US/0898925  
Patent No. 5989820  
GENERAL INFORMATION:  
APPLICANT: Bandman, Olga  
APPLICANT: Lal, Preeti  
APPLICANT: Shah, Puri  
APPLICANT: Corley, Neil C.  
TITLE OF INVENTION: HUMAN ADIPOPHILIN-LIKE PROTEIN  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/989,925  
FILING DATE: Herewith  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0440 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650-855-0555  
TELEFAX: 650-845-4166  
TELEX:



;; INFORMATION FOR SEQ ID NO: 3:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 437 amino acids  
;; TYPE: amino acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; IMMEDIATE SOURCE:  
;; LIBRARY: GenBank  
;; CLONE: 1806040  
US-08-989-925-3

Query Match 45.8%; Score 22; DB 2; Length 437;  
Best Local Similarity 27.8%; Pred. No. 2e+02;  
Matches 5; Conservative 3; Mismatches 10; Indels 0; Gaps 0;

Db 2 EEKXXLXXXXXXXXVXN 19  
91 EERLPILNQPSTQIVAN 108

RESULT 5  
US-08-973-462-9  
Sequence 9, Application US/08973462B  
Patent No. 6191270

;; GENERAL INFORMATION:  
;; APPLICANT: DRUILHE, PIERRE  
;; APPLICANT: DAUBERSIES, PIERRE  
;; TITLE OF INVENTION: MALARIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES  
;; FILE REFERENCE: 0660-0125-0 PCT  
;; CURRENT APPLICATION NUMBER: US/08/973,462B  
;; EARLIER FILING DATE: 1998-02-06  
;; EARLIER APPLICATION NUMBER: PCT/FR96/00894  
;; EARLIER FILING DATE: 1996-06-12  
;; EARLIER APPLICATION NUMBER: FR 95/07007  
;; NUMBER OF SEQ ID NOS: 29  
;; SOFTWARE: PatentIn Ver. 2.0  
;; SEQ ID NO 9  
;; LENGTH: 630  
;; TYPE: PPT  
;; ORGANISM: Artificial Sequence  
;; FEATURE:  
;; OTHER INFORMATION: Description of Artificial Sequence: polypeptide  
US-08-973-462-9

Query Match 43.8%; Score 21; DB 4; Length 630;  
Best Local Similarity 29.4%; Pred. No. 5.3e+02;  
Matches 5; Conservative 2; Mismatches 10; Indels 0; Gaps 0;

Db 2 EEKXXLXXXXXXXXVXN 18  
593 EEKVDLENVVSILDN 609

RESULT 6  
US-08-973-462-8  
Sequence 8, Application US/08973462B  
Patent No. 6191270  
;; GENERAL INFORMATION:  
;; APPLICANT: DRUILHE, PIERRE  
;; APPLICANT: DAUBERSIES, PIERRE  
;; TITLE OF INVENTION: MALARIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES  
;; FILE REFERENCE: 0660-0125-0 PCT  
;; CURRENT APPLICATION NUMBER: US/08/973,462B  
;; EARLIER FILING DATE: 1998-02-06  
;; EARLIER APPLICATION NUMBER: PCT/FR96/00894  
;; EARLIER FILING DATE: 1996-06-12  
;; EARLIER APPLICATION NUMBER: FR 95/07007  
;; EARLIER FILING DATE: 1995-06-13  
;; NUMBER OF SEQ ID NOS: 29  
;; SOFTWARE: PatentIn Ver. 2.0

;; SEQ ID NO 8  
;; LENGTH: 1786  
;; TYPE: PPT  
;; ORGANISM: Artificial Sequence  
;; FEATURE:  
;; OTHER INFORMATION: Description of Artificial Sequence: polypeptide  
US-08-973-462-8

Query Match 43.8%; Score 21; DB 4; Length 1786;  
Best Local Similarity 29.4%; Pred. No. 1.5e+03;  
Matches 5; Conservative 2; Mismatches 10; Indels 0; Gaps 0;

Db 2 EEKXXLXXXXXXXXVXN 18  
1038 EEKVDLENVVSILDN 1054

RESULT 7  
US-08-276-852-86  
Sequence 86, Application US/08276852  
Patent No. 5652138

;; GENERAL INFORMATION:  
;; APPLICANT: Burton, Dennis R  
;; APPLICANT: Barbas, Carlos F  
;; APPLICANT: Lerner, Richard A  
;; TITLE OF INVENTION: HUMAN NEUTRALIZING MONOCLONAL ANTIBODIES  
;; NUMBER OF SEQUENCES: 170  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: The Scripps Research Institute, Office of  
;; STREET: 10666 NO. 5652138th Torrey Pines Road, Suite 220,  
;; CITY: La Jolla  
;; STATE: CA  
;; COUNTRY: USA  
;; ZIP: 92037  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0, Version #1.25  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/276,852  
;; FILING DATE: 18-JUL-1994  
;; CLASSIFICATION: 514  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/178,302  
;; FILING DATE: 30-SEP-1993  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 07/954,148  
;; FILING DATE: 30-SEP-1992  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Fitting, Thomas  
;; REGISTRATION NUMBER: 34,163  
;; REFERENCE/DOCKET NUMBER: SCR1452P  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 619-554-2937  
;; TELEFAX: 619-554-6312  
;; INFORMATION FOR SEQ ID NO: 86:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 108 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
US-08-276-852-86

Query Match 41.7%; Score 20; DB 1; Length 108;  
Best Local Similarity 25.0%; Pred. No. 1.7e+02;  
Matches 4; Conservative 2; Mismatches 10; Indels 0; Gaps 0;

OY 3 EKKXLLXXXXXXVYN 18  
1: 1: 1:  
Db 15 ERATLSCRASQSVISN 30

## RESULT 8

US-08-899-575-86

Sequence 86, Application US/08899575

Patent No. 5770440

GENERAL INFORMATION:

APPLICANT: Burton, Dennis R

APPLICANT: Barbas, Carlos F

APPLICANT: Lerner, Richard A

TITLE OF INVENTION: HUMAN NEUTRALIZING MONOCLONAL ANTIBODIES

TITLE OF INVENTION: TO HUMAN IMMUNODEFICIENCY VIRUS

NUMBER OF SEQUENCES: 170

CORRESPONDENCE ADDRESSES:

ADDRESSEE: The Scripps Research Institute, Office of

ADDRESSEE: Patent Counsel

STREET: 10666 No. 5770440th Torrey Pines Road, Suite 220,

STREET: Mail Drop TPC8

CITY: La Jolla

STATE: CA

COUNTRY: USA

ZIP: 92037

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/899,575

FILING DATE: 24-JUL-1997

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/276,852

FILING DATE: 18-JUL-1994

APPLICATION NUMBER: US 08/178,302

FILING DATE: 30-SEP-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/954,148

FILING DATE: 30-SEP-1992

ATTORNEY/AGENT INFORMATION:

NAME: Fitting, Thomas

REGISTRATION NUMBER: 34,163

REFERENCE/DOCKET NUMBER: SCR1452P

TELECOMMUNICATION INFORMATION:

TELEPHONE: 619-554-2937

TELEFAX: 619-554-6312

INFORMATION FOR SEQ ID NO: 86:

SEQUENCE CHARACTERISTICS:

LENGTH: 108 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-899-575-86

Query Match 41.7%; Score 20; DB 1; Length 108;

Best Local Similarity 25.0%; Pred. No. 1.7e+02;

Matches 4; Conservative 2; Mismatches 10; Indels 0; Gaps 0;

OY 3 EKKXLLXXXXXXVYN 18  
1: 1: 1:  
Db 15 ERATLSCRASQSVISN 30

## RESULT 9

US-08-899-575-86

Sequence 86, Application US/08899575

Patent No. 5804440

GENERAL INFORMATION:

APPLICANT: Burton, Dennis R

APPLICANT: Barbas, Carlos F  
APPLICANT: Lerner, Richard A  
TITLE OF INVENTION: HUMAN NEUTRALIZING MONOCLONAL ANTIBODIES  
TITLE OF INVENTION: TO HUMAN IMMUNODEFICIENCY VIRUS  
NUMBER OF SEQUENCES: 170  
CORRESPONDENCE ADDRESSES:

ADDRESSEE: The Scripps Research Institute, Office of

ADDRESSEE: Patent Counsel

STREET: 10666 No. 5804440th Torrey Pines Road, Suite 220,

STREET: Mail Drop TPC8

CITY: La Jolla

STATE: CA

COUNTRY: USA

ZIP: 92037

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/899,575

FILING DATE: 24-JUL-1997

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/276,852

FILING DATE: 18-JUL-1994

APPLICATION NUMBER: US 08/178,302

FILING DATE: 30-SEP-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/954,148

FILING DATE: 30-SEP-1992

ATTORNEY/AGENT INFORMATION:

NAME: Fitting, Thomas

REGISTRATION NUMBER: 34,163

REFERENCE/DOCKET NUMBER: SCR1452P

TELECOMMUNICATION INFORMATION:

TELEPHONE: 619-554-2937

TELEFAX: 619-554-6312

INFORMATION FOR SEQ ID NO: 86:

SEQUENCE CHARACTERISTICS:

LENGTH: 108 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-899-575-86

Query Match 41.7%; Score 20; DB 1; Length 108;

Best Local Similarity 25.0%; Pred. No. 1.7e+02;

Matches 4; Conservative 2; Mismatches 10; Indels 0; Gaps 0;

OY 3 EKKXLLXXXXXXVYN 18  
1: 1: 1:  
Db 15 ERATLSCRASQSVISN 30

## RESULT 10

PCT-US95-08743-86

Sequence 86, Application PC/TUS9508743

GENERAL INFORMATION:

APPLICANT:

TITLE OF INVENTION: HUMAN NEUTRALIZING MONOCLONAL ANTIBODIES

TITLE OF INVENTION: TO HUMAN IMMUNODEFICIENCY VIRUS

NUMBER OF SEQUENCES: 170

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US95/08743

FILING DATE: 11-JUL-1995

PRIOR APPLICATION DATA:

```

: APPLICATION NUMBER: US 08/276,852
: FILING DATE: 18-JUL-1994
: INFORMATION FOR SEQ ID NO: 86:
: SEQUENCE CHARACTERISTICS
: LENGTH: 108 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: PCT-US93-08743-86

```

Query Match	41.7%;	Score 20;	DB 5;	Length 108
Best Local Similarity	25.0%;	Pred. No. 1.7e+02;		
Matches	4;	Conservative	2;	Mismatches 10;
				Indels

```
QY      3  E K X X L X X X X X V V X N 18
          | : | | : |
Db      15  E R A T L S C R A S Q S V I S N 30
```

RESULT 11  
S-08-946-329A-63  
Sequence 63, Application US/08946329A  
Patent No. 6057091

: GENERAL INFORMATION:  
 : APPLICANT: Beachy, Philip A.  
 : APPLICANT: Porter, Jeffrey A.  
 : TITLE OF INVENTION: NOVEL HEDGEHOG-DERIVED POLYPEPTIDES  
 : NUMBER OF SEQUENCES: 109  
 : CORRESPONDENCE ADDRESS:  
 : ADDRESSEE: Fish & Richardson P.C.  
 : STREET: 4225 Executive Square, Suite 1400  
 : CITY: LA JOLLA  
 : STATE: CA  
 :

```

:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: Windows 95
: SOFTWARE: FASTSEO for Windows Version 2.0b
:
: CURRENT APPLICATION DATA: 00 00 00 00
:

```

```

1 RECORD DATE: 02 DEC 1994
2
3 ATTORNEY/AGENT INFORMATION:
4
5 NAME: Halle, Lisa A.
6
7 REGISTRATION NUMBER: 38,347
8
9 REFERENCE/DOCKET NUMBER: 07265/140001
10
11 TELECOMMUNICATION INFORMATION:
12
13 TELEPHONE: 619/678-5070
14
15 TELEFAX: 619/678-5099
16
17 INFORMATION FOR SEQ ID NO: 63:
18
19 SEQUENCE CHARACTERISTICS:
20
21 LENGTH: 148 amino acids
22
23 TYPE: amino acid
24
25 TOPOLOGY: linear
26
27 MOLECULE TYPE: peptide
28
29 US-08-946-329A-63

```

Query Match	41.7%	Score 20;	DB 3;	Length 148;
Best Local Similarity	29.4%	Pred. No. 2.3e+02;		
Matches	5;	Conservative	2;	Mismatches 10;
				Indels 0;
				Gaps 0

```
QY      3 EKXXLXXXXXXXXXNA 19
          || | | : :
Db      65 EKITLTAHLFLVLDNS 81
```

RESULT 12  
US-09-475-316A-118  
; Sequence 118, Application US/09475316A  
; Patent No. 6210942  
; GENERAL INFORMATION:  
; INFORMATION: US 601009222 C

1 APPLICANT: LEWIS, NO. 0610947man G.  
 2 APPLICANT: Davin, Laurence B.  
 3 APPLICANT: Dinkova-Koslova, Albena T.  
 4 APPLICANT: Fujita, Masayuki  
 5 APPLICANT: Gang, David R.  
 6 APPLICANT: Sarkanen, Simo  
 7 APPLICANT: Ford, Joshua D.  
 8 TITLE OF INVENTION: RECOMBINANT PINORESINOL/LARICIRESNOL REDUCTASES  
 9 TITLE OF INVENTION: RECOMBINANT DIRIGENT PROTEINS AND METHODS OF USE  
 10 FILE REFERENCE: WSUR-1-13793  
 11 CURRENT APPLICATION NUMBER: US/09/475, 316A

```
; SEQ ID NO 118
;
; LENGTH: 314
; TYPE: PRT
; ORGANISM: Schisandra chinensis
US-09-475-316A-118
```

Query Match	41.7%;	Score 20;	DB 4;	Length 314;
Best Local Similarity	29.4%;	Pred. NO. 4.9e+02;		
Matches	5;	Conservative	2;	Mismatches 10;
				Indels 0;
				Gaps 0

```
QY      2  EEKXLLXXXXXXXXVVXN 18
      .  :||| | :|
Db      181 QEKVLEFGDEAKVIWN 19
```

```

1      RESULT 13
2      US-08-176-427B-10
3      ; Sequence 10, Application US/08176427B
4      ; Patent No. 5789543
5      ; GENERAL INFORMATION:
6      ; APPLICANT: Ingham, Phillip W.
7      ; APPLICANT: McMahon, Andrew P.
8      ; APPLICANT: Tablin, Clifford J.
9      ; TITLE OF INVENTION: Vertebrate Embryonic Pattern-inducing
10     ; TITLE OF INVENTION: Proteins and Uses Related Thereto
11     ; NUMBER OF SEQUENCES: 33
12     ; CORRESPONDENCE ADDRESS:
13     ; ADDRESSEE: LAHIVE & COCKFIELD
14     ; STREET: 60 State Street
15     ; CITY: Boston
16     ; STATE: MA
17     ; COUNTRY: USA
18     ; ZIP: 02109
19     ; COMPUTER READABLE FORM:
20     ; MEDIUM TYPE: Floppy disk
21     ; COMPUTER: IBM PC compatible
22     ; OPERATING SYSTEM: PC-DOS/MS-DOS
23     ; SOFTWARE: ASCII(text)
24     ; CURRENT APPLICATION DATA:
25     ; APPLICATION NUMBER: US/08/176.427B

```



Fri Mar 1 07:28:30 2002

us-09-359-426c-1.ra1

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

# OM protein - protein search, using sw model

Run on: February 28, 2002, 11:33:24 : Search time 23.26 Seconds  
(without alignments)  
62.223 Million cell updates/sec

Title: US-09-359-426c-1

Perfect score: 48  
Sequence: 1 XEKKXXLXXXXXXVXNA 19

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : PIR68:\*  
1: PIR1:\*  
2: PIR2:\*  
3: PIR3:\*  
4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Query Length	DB ID	Description
1	28	58.3	482	2 B83113	catalase PA4236 [1
2	26	54.2	514	2 T48438	hypothetical prote
3	24	50.0	214	2 H69830	conserved hypothet
4	24	50.0	408	2 T32767	hypothetical prote
5	23	47.9	113	2 A81439	hypothetical prote
6	23	47.9	205	2 C83521	probable fibrillar
7	23	47.9	251	2 C86680	prophage pil prote
8	23	47.9	1111	2 T01078	hypothetical prote
9	22	45.8	202	2 T14970	phage lambda-relat
10	22	45.8	266	2 D69776	hypothetical prote
11	22	45.8	387	2 A86302	hypothetical prote
12	22	45.8	484	2 A58663	catalase (EC 1.11.
13	22	45.8	491	2 T16354	hypothetical prote
14	22	45.8	653	2 E86787	hypothetical prote
15	22	45.8	689	2 H70024	sorbitol-6-phospha
16	22	45.8	742	2 T33514	hypothetical prote
17	22	45.8	2194	1 GNNYE7	genome polyprotein
18	22	45.8	3085	2 T00327	polyprotein - infe
19	21	43.8	144	2 F64012	hypothetical prote
20	21	43.8	238	2 D71090	probable arylamon
21	21	43.8	262	2 T33408	hypothetical prote
22	21	43.8	304	2 B86641	hypothetical prote
23	21	43.8	391	2 T16673	hypothetical prote
24	21	43.8	405	2 C71462	hypothetical prote
25	21	43.8	412	2 T05285	farnesyl-diphospha
26	21	43.8	413	2 T44924	probable squalene
27	21	43.8	413	2 E85408	glutamate-1-semial
28	21	43.8	481	2 T07034	thiamin biosynthes
29	21	43.8	515	2 E82267	

30	21	43.8	534	2 E82269	conserved hypothet
31	21	43.8	608	2 S76192	hypothetical prote
32	21	43.8	832	2 T23693	hypothetical prote
33	21	43.8	1429	2 T41699	C2-domain family p
34	21	43.8	1558	2 B71603	RESA-H3 antigen p
35	21	43.8	1683	2 T30885	complement compone
36	21	43.8	1776	1 RRPPTM	genome polyprotein
37	21	43.8	1844	1 RRPPTM	genome polyprotein
38	21	43.8	1844	2 S01956	hypothetical prote
39	21	43.8	1885	2 S19151	hypothetical prote
40	21	43.8	2314	2 T28698	hypothetical prote
41	20	41.7	91	2 T43345	ribosomal protein,
42	20	41.7	97	2 T41233	ribosomal protein,
43	20	41.7	128	2 JC1273	ribosomal protein
44	20	41.7	139	2 C70114	ribosomal protein
45	20	41.7	139	2 B64010	hypothetical prote

## ALIGNMENTS

RESULT 1  
B83113  
catalase PA4236 [imported] - Pseudomonas aeruginosa (strain PA01)  
C:Species: Pseudomonas aeruginosa  
C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
C:Accession: B83113  
R:Stover, C.K.; Pham, X.Q.; Ewlin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.;  
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lartig, K.; L.  
of, J.; Olson, M.V.  
Nature 406, 959-964, 2000  
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa  
A:Reference number: AB2950, NCBI:20437337  
A:Accession: B83113  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-482 <STO>  
A:Cross-references: GB:AE004841; GB:AE004091; NID:g9950451; PIDN:AA07624.1; GSPDB:GN  
C:Genetics:  
A:Experimental source: strain PA01  
C:Superfamily: catalase

Query Match 58.3% Score 28; DB 2; Length 482;  
Best Local Similarity 41.2% Pred. No. 2.7;  
Matches 7; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 2 EKKXXLXXXXXXVXN 18  
Db 2 EKKRLTTAGAPVVDN 18  
RESULT 2  
T48438  
hypothetical protein T32M21.20 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 20-Apr-2000  
C:Accession: T48438  
R:Bevan, M.; Terry, N.; Ardiles, W.; Buysse, C.; Dasseville, R.; De Clerck, R.;  
ewes, H.W.; Rudd, S.; Lemcke, K.; Mayer, K.F.X.  
submitted to the Protein Sequence Database, March 2000  
A:Reference number: 224487  
A:Accession: T48438  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-514 <BEV>  
A:Cross-references: EMBL:AL162875  
A:Experimental source: cultivar Columbia; BAC clone T32M21  
C:Genetics:  
A:Map position: 5  
A:Introns: 37/3; 63/3; 107/3; 130/1; 150/3; 197/2; 239/3; 278/1; 307/3; 335/3; 442/3;  
A:Note: T32M21.20

Query Match 54.2%; Score 26; DB 2; Length 514;  
Best Local Similarity 38.9%; Pred. No. 12;  
Matches 7; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

OY 2 EKKXLLXXXXXXXXXXVXNA 19  
||| : ||  
Db 407 EKKRALESSIAETOVENA 424

RESULT 3  
HE9830  
Conserved hypothetical protein yhfK - Bacillus subtilis

C:Species: Bacillus subtilis  
C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 20-Jun-2000  
C:Accession: H69830

A:Kunst, F.; Ogasawara, N.; Moser, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter  
A:Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Ch  
A:Erlich, S.D.; Emerson, P.T.; Entlan, K.D.; Errington, J.; Fabret, C.; Ferrari, E.  
Nature 390, 249-256, 1997

A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Funo, S.; Galizzi, A.; Gall  
A:lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.  
A:Koester, P.; Konigstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Laidonis,  
A:V.M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetlell  
A:Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon,  
A:Authors: Schleich, S.; Schoeter, R.; Scottone, F.; Sekiguchi, J.; Sekowska, A.; Serot  
A:keuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,  
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasunoto, K.; Yata, K.; Yoshida, K

A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.  
A:Title: The complete genome sequence of the gram-positive bacterium Bacillus subtilis.  
A:Reference number: A69580; MUID:98044033

A:Accession: H69830  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-214 <KUN>

A:Cross-references: GB:299109; GB:AL009126; NID:92633260; PIDN:CAB12866.1; PID:92633362  
A:Experimental source: strain 168  
C:Genetics:  
A:Gene: yhfK

C:Superfamily: hypothetical protein YMR090w

Query Match 50.0%; Score 24; DB 2; Length 214;  
Best Local Similarity 29.4%; Pred. No. 18;  
Matches 5; Conservative 2; Mismatches 10; Indels 0; Gaps 0;

OY 2 EKKXLLXXXXXXXXXXVXN 18  
||| : ||  
Db 36 EOKASLEAAGAVLAN 52

RESULT 4  
T32767  
hypothetical protein F33D11.2 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans  
C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999  
C:Accession: T32767  
R:Sammons, L.; Wohldmann, P.; Mullen, G.  
submitted to the EMBL data library, December 1997

A:Description: The sequence of C. elegans cosmid F33D11.  
A:Reference number: 221222  
A:Accession: T32767

A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-408 <SAM>

A:Cross-references: EMBL:AF039720; PIDN:AAB96696.1; GSPDB:GN00019; CESP:F33D11.2  
A:Experimental source: strain Bristol N2; clone F33D11  
C:Genetics:  
A:Gene: CESP:F33D11.2  
A:Map position: 1  
A:Insertions: 11/3; 54/2; 93/3; 188/3; 268/3; 385/3

Query Match 50.0%; Score 24; DB 2; Length 408;  
Best Local Similarity 33.3%; Pred. No. 37;  
Matches 6; Conservative 1; Mismatches 11; Indels 0; Gaps 0;

OY 2 EKKXLLXXXXXXXXXXVXNA 19  
||| : ||  
Db 311 EKKALSIKIAVAVLANA 328

RESULT 5  
AB1439  
hypothetical protein Cj0202c [imported] - Campylobacter jejuni (strain NCTC 11168)

C:Species: Campylobacter jejuni  
C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 28-Jul-2000  
C:Accession: AB1439

R:Parkhill, J.; Wren, B.W.; Mungall, K.; Kelley, J.M.; Churcher, C.; Basham, D.; Chl  
C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVleet, A.; Whitehead, S.; Ba  
Nature 403, 665-668, 2000

A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals  
A:Reference number: AB1250; MUID:20150912  
A:Accession: AB1439

A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-113 <PAR>

A:Cross-references: GB:AL139074; GB:AL111168; NID:96967505; PIDN:CAB72685.1; PID:9696  
A:Experimental source: serotype O2, strain NCTC 11168  
C:Genetics:  
A:Gene: Cj0202c

C:Superfamily: Campylobacter jejuni hypothetical protein Cj0202c

Query Match 47.9%; Score 23; DB 2; Length 113;  
Best Local Similarity 33.3%; Pred. No. 19;  
Matches 6; Conservative 1; Mismatches 11; Indels 0; Gaps 0;

OY 2 EKKXLLXXXXXXXXXXVXNA 19  
||| : ||  
Db 40 EKFVSDKITQVAVENA 57

RESULT 6  
C83521  
probable fibrial subunit protein PA0992 [imported] - Pseudomonas aeruginosa (strain

C:Species: Pseudomonas aeruginosa  
C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
C:Accession: C83521  
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warriner, P.; Hickey, M.J.;  
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L  
Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000

A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa  
A:Reference number: A82950; MUID:20437337  
A:Accession: C83521

A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-205 <STO>  
A:Cross-references: GB:AF004532; GB:AF004091; NID:99946896; PIDN:AA04381.1; GSPDB:GN  
A:Experimental source: strain PA01  
C:Genetics:  
A:Gene: PA0992

Query Match 47.9%; Score 23; DB 2; Length 205;  
Best Local Similarity 37.5%; Pred. No. 35;  
Matches 6; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

OY 4 KXXLLXXXXXXXXXXVXNA 19  
||| : ||  
Db 6 KSIIAANVASLVVNA 21



RESULT 7  
C86680  
Prophage p11 protein 08 [imported] - *Lactococcus lactis* subsp. *lactis* (strain IL1403)  
C:Species: *Lactococcus lactis* subsp. *lactis*  
C:Date: 23-Mar-2001 #sequence\_revision 23-Mar-2001 #text\_change 23-Mar-2001  
C:Accession: C86680  
R:Holotin, A.; Wincker, P.; Mueger, S.; Jallion, O.; Malarme, K.; Weissbach, J.; Ehrlich  
Genome Res. In press, 2001  
A:Title: The complete genome sequence of the lactic acid bacterium.  
A:Reference number: A86625  
A:Accession: C86680  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-251 <STO>  
A:Cross-references: GB:AE005176; NID:g12723321; PIDN:AAK04541.1; GSPDB:GN00146  
C:Genetics:  
A:Gene: p1108

Query Match 47.9%; Score 23; DB 2; Length 251;  
Best Local Similarity 35.3%; Pred. No. 44;  
Matches 6; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

OY 2 EEXXXLXXXXXXVXN 18  
Db 133 EEAQLALNSTLAVEN 149

RESULT 8  
T01078  
hypothetical protein T10P11.2.2 - *Arabidopsis thaliana*  
C:Species: *Arabidopsis thaliana* (mouse-ear cross)  
C:Date: 12-Feb-1999 #sequence\_revision 12-Feb-1999 #text\_change 24-Mar-1999  
C:Accession: T01078  
R:Kaplan, N.; Johnson, D.; Schutz, K.; Gnoj, L.; Hoffman, J.; Tili, S.; de la Bastide, M.  
M.; Martienssen, R.; Chen, E.Y.; Willson, R.; McComble, W.R.  
Submitted to the EMBL Data Library, November 1998  
A:Description: Sequence of A. thaliana BAC T10P11 from chromosome IV.  
A:Reference number: Z14248  
A:Accession: T01078  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1111 <KAP>  
A:Cross-references: EMBL:AC002330; NID:g2262135; PID:g3892059  
A:Experimental source: cultivar Columbia  
C:Genetics:  
A:Map position: 4  
A:Introns: 35/1; 795/3  
A:Note: T10P11.2.2

Query Match 47.9%; Score 23; DB 2; Length 1111;  
Best Local Similarity 35.3%; Pred. No. 2.2e+02;  
Matches 6; Conservative 1; Mismatches 10; Indels 0; Gaps 0;

OY 2 EEXXXLXXXXXXVXN 18  
Db 597 EEXSCLISEQNHVNIEN 613

RESULT 9  
T14970  
phage lambda-related tail assembly protein G - *Yersinia pestis* plasmid pMT1  
C:Species: *Yersinia pestis*  
C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 01-Dec-2000  
C:Accession: T14970; T14649  
R:Findler, L.E.; Plano, G.V.; Burland, V.; Mayhew, G.F.; Blattner, F.R.  
Infect. Immun. 66, 5731-5742, 1998  
A:Title: Complete DNA sequence and detailed analysis of the *Yersinia pestis* KIMS plasmid  
A:Reference number: Z18268; M0ID:99043898  
A:Accession: T14970  
A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA  
A:Residues: 1-202 <LIN>  
A:Cross-references: EMBL:AF074611; NID:g3883003; PID:g3883053; PIDN:AC82713.1  
R:Hu, P.; Elliott, J.; McCreedy, P.; Skowronski, E.; Barnes, J.; Kobayashi, A.; Carra  
Submitted to the EMBL Data Library, March 1998  
A:Description: Structural organization of virulence determinants in three *Yersinia pe*  
A:Reference number: Z18168  
A:Accession: T14649  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 5-202 <HUP>  
A:Cross-references: EMBL:AF053947; NID:g2996286; PID:g2996310; PIDN:AAC13190.1  
C:Genetics:  
A:Gene: y1053  
A:Gene: plasmid pMT1  
C:Superfamily: phage T4 tail fiber assembly protein gp38

Query Match 45.8%; Score 22; DB 2; Length 202;  
Best Local Similarity 35.3%; Pred. No. 70;  
Matches 6; Conservative 1; Mismatches 10; Indels 0; Gaps 0;

OY 2 EEXXXLXXXXXXVXN 18  
Db 159 EKETLVALKKRYVLN 175

RESULT 10  
D69776  
hypothetical protein yddk - *Bacillus subtilis*  
C:Species: *Bacillus subtilis*  
C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 20-Jun-2000  
C:Accession: D69776  
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber  
C.; Bron, S.; Brouillette, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;  
A.; Ehlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari,  
Nature 390, 249-256, 1997  
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galluzzi, A.; Gal  
lechi, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holtsappel, S.; Hosono, S.; Hullo, M  
Koetler, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino  
A.; Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mu  
Y., M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete  
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scari  
A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Se  
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terstra, P.; Tognoni, A.; Tosato, V.; Uchiya  
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida  
A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.  
A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*  
A:Reference number: A69580; M0ID:98044033  
A:Accession: D69776  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-266 <KUN>  
A:Cross-references: GB:Z99106; GB:AL009126; NID:g2632653; PIDN:CA812307.1; PID:g26328  
A:Experimental source: strain 168  
C:Genetics:  
A:Gene: yddk  
C:Superfamily: *Bacillus subtilis* hypothetical protein yddk

Query Match 45.8%; Score 22; DB 2; Length 266;  
Best Local Similarity 27.8%; Pred. No. 94;  
Matches 5; Conservative 3; Mismatches 10; Indels 0; Gaps 0;

OY 2 EEXXXLXXXXXXVXN 19  
Db 113 EKKFISHSSKRIYCN 130

RESULT 11  
A86302  
hypothetical protein AAG10814.1 [imported] - *Arabidopsis thaliana*  
C:Species: *Arabidopsis thaliana* (mouse-ear cross)

C>Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
 C:Accession: A86302  
 R:Theologidis, A.; Becker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, C.M.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; Jensen, N.F.; Hughes, B.; Hulzar, L.  
 Nature 408, 816-820, 2000  
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
 A:Authors: Salberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, Ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
 A>Title: Sequence and analysis of chromosome 1 of the plant *Arabidopsis*.  
 A:Reference number: A86141; MUID:21016719  
 A:Accession: A86302  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-387 <STO>  
 A:Cross-references: GB:AE005172; NID:g9989051; PIDN:AG10814.1; GSPDB:GN00141  
 A:Genetics:  
 A:Map position: 1

Query Match 45.8%; Score 22; DB 2; Length 387;  
 Best Local Similarity 29.4%; Pred. No. 1.4e+02;  
 Matches 5; Conservative 1; Mismatches 11; Indels 0; Gaps 0;

OY 2 EKKXILXXXXXXVXN 18  
 I: I  
 Db 80 EDKSSTKPASSSTVSN 96

RESULT 12  
 A58663  
 Catalase (EC 1.11.1.6) [validated] - Proteus mirabilis  
 C:Species: Proteus mirabilis  
 A:Variety: strain pr, peroxide resistant  
 C>Date: 19-Nov-1997 #sequence\_revision 21-Nov-1997 #text\_change 15-Sep-2000  
 C:Accession: A58663; B58663  
 R:Burzy, A.; Bracchi, V.; Sterlides, R.; Chroboczek, J.; Thibault, P.; Gagnon, J.; Jouve

J. Protein Chem. 14, 59-72, 1995  
 A>Title: Complete amino acid sequence of Proteus mirabilis PR catalase. Occurrence of a  
 A:Reference number: A58663; MUID:95305957  
 A:Accession: A58663  
 A:Molecule type: protein  
 A:Residues: 1-484 <BUZ1>  
 A:Accession: B58663  
 A:Molecule type: DNA  
 A:Residues: 1-305, 'AE' <BUZ2>

R:Gouet, P.; Jouve, H.M.; Dideberg, O.

submitted to the Brookhaven Protein Data Bank, June 1996

A:Reference number: A67899; PDB:2CAE

A:Contents: annotation; X-ray crystallography, 2.2 angstroms, residues 3-475

R:Gouet, P.; Jouve, H.M.; Hajdu, J.

submitted to the Brookhaven Protein Data Bank, June 1996

A:Reference number: A67900; PDB:2CAE

A:Contents: annotation; X-ray crystallography, 2.7 angstroms, residues 3-475

A>Note: compound I

R:Gouet, P.; Jouve, H.M.; Hajdu, J.

submitted to the Brookhaven Protein Data Bank, June 1996

A:Reference number: A67901; PDB:2CAE

A:Contents: annotation; X-ray crystallography, 2.7 angstroms, residues 3-475

A>Note: compound II, dithiothreitol reduced compound I

R:Gouet, P.; Jouve, H.M.; Dideberg, O.

submitted to the Brookhaven Protein Data Bank, July 1996

A:Reference number: A67902; PDB:2CAE

A:Contents: annotation; X-ray crystallography, 2.7 angstroms, residues 3-475

A>Note: native Fe(III) with NADPH

R:Gouet, P.; Jouve, H.M.; Dideberg, O.

J. Mol. Biol. 249, 933-954, 1995

A>Title: Crystal structure of Proteus mirabilis PR catalase with and without bound NADPH  
 A:Reference number: A58664; MUID:95311317  
 A:Contents: annotation; X-ray crystallography, 2.2 angstroms  
 A:Complex: homotrimer

C:Function:  
 A:Description: catalyzes the conversion of two of molecules of hydrogen peroxide to t  
 A>Note: this enzyme has a tightly bound NADPH cofactor  
 C:Superfamily: catalase  
 C:Keywords: Chromoprotein; heme; homotrimer; iron; metalloprotein; NADP; oxidoreduc  
 F:53/Modified site: methionine sulfone (Met) #status experimental  
 F:54.93.127/Active site: His, Ser, Asn #status predicted  
 F:337/Binding site: heme iron (Tyr) (axial ligand) #status experimental

Query Match 45.8%; Score 22; DB 2; Length 484;  
 Best Local Similarity 37.5%; Pred. No. 1.8e+02;  
 Matches 6; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

OY 3 EKKXILXXXXXXVXN 18  
 I: I  
 Db 2 EKKKLTAAAGAPVVDN 17

RESULT 13  
 T16354  
 Hypothetical protein F42G9.1 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C>Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 20-Sep-1999  
 C:Accession: T16354  
 R:Rajch, A.

submitted to the EMBL Data Library, March 1996  
 A:Description: The sequence of C. elegans cosmid F42G9.  
 A:Reference number: T16354  
 A:Accession: T16354  
 A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA  
 A:Residues: 1-491 <TAI>  
 A:Cross-references: EMBL:U00051; NID:g1216305; PIDN:Q1216310; PIDN:AA91358.1; CESP:F4

A:Experimental source: strain Bristol N2  
 C:Genetics:  
 A:Gene: CESP:F42G9.1  
 A:introns: 37/3; 254/2; 364/2; 406/2

Query Match 45.8%; Score 22; DB 2; Length 491;  
 Best Local Similarity 33.3%; Pred. No. 1.8e+02;  
 Matches 6; Conservative 1; Mismatches 11; Indels 0; Gaps 0;

OY 2 EKKXILXXXXXXVXNA 19  
 I: I  
 Db 193 EKKKNEEDASAEVIEA 210

RESULT 14  
 E86787  
 Hypothetical protein [imported] - Lactococcus lactis subsp. lactis (strain IL1403)

C:Species: Lactococcus lactis subsp. lactis

C>Date: 23-Mar-2001 #sequence\_revision 23-Mar-2001 #text\_change 23-Mar-2001

C:Accession: E86787

R:Boillot, A.; Winkler, P.; Manger, S.; Jallion, O.; Malarme, K.; Weissenbach, J.; Eh

Genome Res. in press, 2001

A>Title: The complete genome sequence of the lactic acid bacterium.

A:Reference number: A86625

A:Accession: E86787

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-653 <STO>

A:Cross-references: GB:AE005176; NID:g12724279; PIDN:AAK05399.1; GSPDB:GN00146

A:Experimental source: strain IL1403

C:Genetics:  
 A:Gene: yndf

Query Match 45.8%; Score 22; DB 2; Length 653;  
 Best Local Similarity 29.4%; Pred. No. 2.5e+02;  
 Matches 5; Conservative 2; Mismatches 10; Indels 0; Gaps 0;

OY 3 EKKXLLXXXXXXXXXNA 19  
 || | : ||  
 DB 534 EKKELTGYTFKNICNA 550

## RESULT 15

H70024

sorbitol-6-phosphate 2-dehydrogenase homolog yuxG - *Bacillus subtilis*C:Species: *Bacillus subtilis*

C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 15-Oct-1999

C:Accession: H70024; E54078

R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter

C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Ch

A.; Ehrlich, S.D.; Emerson, P.T.; Enliam, K.D.; Errington, J.; Fabret, C.; Ferrari, E.

Nature 390, 249-256, 1997

A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Gallizzi, A.; Gall

lechi, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hulio, M.F.

Koester, P.; Konlingstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,

A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel

Y. M.; Ogawa, K.; Ogilwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle

Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon,

A:Authors: Schleich, S.; Schroeder, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seron

euch, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,

A:Authors: Yoshikawa, H.F.; Zumstein, E.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.

A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.

A:Reference number: A69580; MUID:98044033

A:Accession: H70024

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-689 &lt;KUN&gt;

A:Cross-references: GB:299119; GB:299120; GB:AL009126; NID:92635613; PIDN:CAB1511.1; PI

A:Experimental source: strain 168

R:Hanlon, D.W.; Ordal, G.W.

J. Biol. Chem. 269, 14038-14046, 1994

A:Title: Cloning and characterization of genes encoding methyl-accepting chemotaxis pro

A:Reference number: A54078; MUID:94245722

A:Accession: E54078

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-31 &lt;HAN&gt;

A:Cross-references: GB:L29189

C:Genetics:

A:Gene: yuxG

C:Superfamily: short-chain alcohol dehydrogenase homology

F:428-610/Domain: short-chain alcohol dehydrogenase homology &lt;SADH&gt;

## Query Match

Best Local Similarity 45.8%; Score 22; DB 2; Length 689;

Matches 6; Conservative 1; Mismatches 10; Indels 0; Gaps 0;

OY 3 EKKXLLXXXXXXXXXNA 19

|| | : ||

DB 498 ERAALAYGIDIVVNA 514

Search completed: February 28, 2002, 11:35:08  
 Job time: 104 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 28, 2002, 11:34:44 ; Search time 12.98 Seconds  
(without alignments)  
53.670 Million cell updates/sec

Title: US-09-359-426C-1

Perfect score: 48  
Sequence: 1 XEKKXXLXXXXXXVXNA 19

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues  
Total number of hits satisfying chosen parameters: 100059

Minimum DB seq Length: 0  
Maximum DB seq Length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database: SWISSPROT\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	28	58.3	482	1	CATA_PSEAE
2	25	52.1	450	1	ADFP_BOVIN
3	22	45.8	437	1	ADFP_HUMAN
4	22	45.8	484	1	CATP_PROMI
5	22	45.8	491	1	P2C1_CAEEL
6	22	45.8	689	1	YUUG_BACSU
7	22	45.8	2194	1	POLG_HE701
8	21	43.8	144	1	Y850_HAEIN
9	21	43.8	481	1	GSA_LYCES
10	21	43.8	692	1	GYRH_BARBA
11	21	43.8	1439	1	YC31_SCHPO
12	21	43.8	1776	1	POLR_OTMIV
13	21	43.8	1844	1	POLR_TYMA
14	21	43.8	1844	1	POLR_TYMA
15	21	43.8	1844	1	POLR_TYMA
16	20	41.7	128	1	RL7_STRAT
17	20	41.7	139	1	RS6_BORBU
18	20	41.7	139	1	Y589_HAEIN
19	20	41.7	154	1	RS2_SULAC
20	20	41.7	174	1	STPA_STAVU
21	20	41.7	240	1	RNC_MYCTU
22	20	41.7	251	1	GLPR_PSEAE
23	20	41.7	259	1	YGFM_ECOLI
24	20	41.7	305	1	E2BA_HUMAN
25	20	41.7	330	1	RLX3_STAVU
26	20	41.7	336	1	YDGC_SCHPO
27	20	41.7	356	1	SHH_BRARE
28	20	41.7	471	1	MEHF_BACSU
29	20	41.7	478	1	GSA_TOBAC
30	20	41.7	519	1	Y771_CAEEL
31	20	41.7	557	1	CGAL_YEAST
32	20	41.7	692	1	Y957_HUMAN
33	20	41.7	737	1	AMY1_AEDAE

34	20	41.7	778	1	TRKA_CHICK
35	20	41.7	809	1	PAT3_CAEEL
36	20	41.7	950	1	MIC1_YEAST
37	20	41.7	1434	1	RPOB_UREPA
38	20	41.7	1690	1	RPOC_THEMA
39	20	41.7	2195	1	POLG_ECOLI
40	20	41.7	3068	1	POLG_PEMVC
41	19	39.6	92	1	RL19_PICAB
42	19	39.6	92	1	RL19_PICAB
43	19	39.6	125	1	RL7_LTBAF
44	19	39.6	132	1	Y055_ARCFU
45	19	39.6	134	1	CLX2_HUMAN

## ALIGNMENTS

```

RESULT 1
CATA_PSEAE STANDARD: PRT: 482 AA.
ID CATA_PSEAE
AC 052762:2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE CATALASE (EC 1.11.1.6).
GN KATA OR PA4236.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FR01.
RA MEDLINE=99296583; PubMed=10368148;
RA Ma J.-F., Ochser U.A., Klotz M.G., Nanayakkara V.K., Howell M.L.,
RA Johnson Z., Posey J.E., Vasil M.L., Monaco J.J., Hasset D.J.,
RA "Bacterioferitin A modulates catalase A (Kata) activity and
RT resistance to hydrogen peroxide in Pseudomonas aeruginosa."
RL J. Bacteriol. 181:3730-3742(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RA MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-O.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltry L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.L., Coulter D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Smith K.A., Spencer D.H., Hong G.K.-S., Lory S., Olson M.V.,
RA Reizer J., Sater M.H., Hancock R.E.W., Lory S., Olson M.V.,
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen."
RL Nature 406:959-964(2000).
CC -1- FUNCTION: DECOMPOSES HYDROGEN PEROXIDE IN WATER AND OXYGEN; SERVES
CC TO PROTECT CELLS FROM THE TOXIC EFFECTS OF HYDROGEN PEROXIDE.
CC -1- CATALYTIC ACTIVITY: 2 H(2)O(2) = O(2) + 2 H(2)O.
CC -1- COFACTOR: HEME GROUP.
CC -1- ENZYME REGULATION: BY PEROXIDE AND BFR-BOUND IRON.
CC -1- SUBUNIT: HETEROMULTIMER. POSSIBLY AN ALPHA(2)BETA-HETEROTRIMER
CC WHERE THE ALPHA SUBUNIT IS A 56 KDA PROTEIN AND THE BETA SUBUNIT A
CC 45 KDA PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE CATALASE FAMILY.
CC THIS SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: AF047025; AAC03118.1; -
CC EMBL: AF004841; AAC07624.1; -
CC InterPro: IPR002226; Catalase.

```



RT "Crystal structure of Proteus mirabilis PR catalase with and without  
bound NADPH.";  
RL J. Mol. Biol. 249:933-954(1995).  
CC -1- FUNCTION: DECOMPOSES HYDROGEN PEROXIDE IN WATER AND OXYGEN; SERVES  
TO PROTECT CELLS FROM THE TOXIC EFFECTS OF HYDROGEN PEROXIDE.  
CC -1- CATALYTIC ACTIVITY: 2 H(2)O(2) = O(2) + 2 H(2)O.  
CC -1- COFACTOR: HEME GROUP AND NADP.  
CC -1- SUBUNIT: HOMOTETRAMER.  
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.  
CC -1- MASS SPECTROMETRY: MW=55643; MW\_ERR=5; METHOD=ELECTROSPRAY.  
CC -1- SIMILARITY: BELONGS TO THE CATALASE FAMILY.  
DR PDB: 2CAE; 07-DEC-96.  
DR PDB: 2CAG; 07-DEC-96.  
DR PDB: 2CAH; 11-JAN-97.  
DR InterPro: IPR002226; Catalase.  
DR Pfam: PF00199; catalase.1.  
DR PRINTS: PR00067; CATALASE.  
DR ProDom: PD000510; Catalase.1.  
DR PROSITE: PS00437; CATALASE\_1; 1.  
DR OXidoreductase: Peroxidase; Iron; Heme; Hydrogen peroxide; NADP;  
3D-structure.  
FT MOD\_RES 53 53 METHIONINE SULFONE.  
FT ACT\_SITE 127 127  
FT ACT\_SITE 337 337  
FT BINDING 337 337 PROXIMAL HEME LIGAND.  
SQ SEQUENCE 484 AA; 55614 MW; ADC25FCB41F5C50 CRC64;

Query Match 45.8%; Score 22; DB 1; Length 484;  
Best Local Similarity 37.5%; Pred. No. 99;  
Matches 6; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 3 EKKXXLXXXXXXVYXN 18  
DB 2 EKKKLTAAAGAPVVDN 17

RESULT 5  
P2C1\_GAEEL STANDARD; PRT; 491 AA.  
AC P49595;  
DT 01-FEB-1996 (Rel. 33, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE PROBABLE PROTEIN PHOSPHATASE 2C P42G9.1 (EC 3.1.3.16) (PP2C).  
GN F42G9.1.  
OS Caenorhabditis elegans.  
NC Eukaryote; Metazoa; Nematoda; Chromadorea; Rhabditidae; Rhabditoidea;  
NC Rhabditidae; Pelodermatidae; Caenorhabditis.  
NCBI\_TaxID=6239;  
RN [1]  
RN SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RA Taich A., Waterston R.;  
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.  
CC -1- CATALYTIC ACTIVITY: A PHOSPHOPROTEIN + H(2)O = A PROTEIN +  
ORPHOSPHATE (THIS ENZYME IS SERINE/THREONINE SPECIFIC).  
CC -1- COFACTOR: BINDS TWO MAGNESIUM OR MANGANESE IONS (BY SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO THE PP2C FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
DR EMBL: U00051; AAA91358.1; -  
DR HSSP: P35813; IAG0.  
DR WormPep: F42G9.1; CE07231.

DR InterPro: IPR002222; PP2C.  
DR InterPro: IPR003589; PP2C\_catalytic.  
DR InterPro: IPR001932; PP2C\_domain.  
DR InterPro: IPR003588; PP2C\_slg.  
DR Pfam: PF00481; PP2C; 2.  
DR SMART: SM00332; PP2C; 1.  
DR SMART: SM00331; PP2C-SIG; 1.  
DR PROSITE: PS01032; PP2C; 1.  
KW Hypothetical protein; Hydrolase; Magnesium; Manganese.  
FT METAL 37 37  
FT METAL 38 38 MANGANESE 1 (BY SIMILARITY).  
FT METAL 57 57 MANGANESE 1 (BY SIMILARITY).  
FT METAL 428 428 MANGANESE 1 AND 2 (BY SIMILARITY).  
FT METAL 477 477 MANGANESE 2 (BY SIMILARITY).  
FT METAL 477 477 MANGANESE 2 (BY SIMILARITY).  
SQ SEQUENCE 491 AA; 53141 MW; F110D12E343953D6 CRC64;

Query Match 45.8%; Score 22; DB 1; Length 491;  
Best Local Similarity 33.3%; Pred. No. 16+02;  
Matches 6; Conservative 1; Mismatches 11; Indels 0; Gaps 0;

QY 2 EKKXXLXXXXXXVYXNA 19  
DB 193 EKKKEDASAEVYIENA 210

RESULT 6  
YUXG\_BACSU STANDARD; PRT; 689 AA.  
AC P40747;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE HYPOTHETICAL OXIDOREDUCTASE IN GBSA-TLPB INTERGENIC REGION  
DE (EC 1.-.-.-) (ORF2).  
GN YUXG OR YULA.  
OS Bacillus subtilis.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
OC Bacillus/Staphylococcus group; Bacillus.  
OX NCBI\_TaxID=1423;  
RN [1]  
RN SEQUENCE FROM N.A.  
RC STRAIN=168;  
RA Oudega B., Koningsleyn G., Feger G., Pohl T.;  
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RN SEQUENCE OF 1-49 FROM N.A.  
RC STRAIN=168 / O11085;  
RX MEDLINE=94245722; Pubmed=8186684;  
RA Hannon D.W., Ordal G.W.;  
RT "Cloning and characterization of genes encoding methyl-accepting  
chemotaxis proteins in Bacillus subtilis.";  
J. Biol. Chem. 269:14038-14046(1994).  
CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASIS/REDUCTASES  
(SDR) FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
DR EMBL: Z93938; CAB07952.1; -  
DR EMBL: L29189; AAA20558.2; -  
DR EMBL: Z99119; CAB15100.1; -  
DR EMBL: Z99120; CAB15111.1; -  
DR HSSP: P25529; IAH1.  
DR Subtilist; BG10947; YUXG.  
DR InterPro: IPR002198; ADH\_short.  
DR Pfam: PF00106; adh\_short.1.  
DR PROSITE: PS00061; ADH\_SHORT; 1.

KW Hypothetical protein; Oxidoreductase; Complete proteome.  
 FT ACT\_SITE 579 579 BY SIMILARITY.  
 SQ SEQUENCE 689 AA; 76020 MW; 88A82FDC3EEDD0FD CRC64;

Query Match 45.8%; Score 22; DB 1; Length 689;  
 Best Local Similarity 35.3%; Pred. No. 1.4e+02;  
 Matches 6; Conservative 1; Mismatches 10; Indels 0; Gaps 0;

QY 3 EKKXLLXXXXXXVYVXNA 19  
 | | | | |  
 Db 498 ERALAVGIDIVVNA 514

RESULT 7  
 POLG\_HE701 STANDARD: PRT: 2194 AA.  
 P32537:  
 01-OCT-1993 (Rel. 32, Last sequence update)  
 20-NOV-1995 (Rel. 40, Last annotation update)  
 DE GENOME POLYPEPTIDE [CONTAINS: COAT PROTEINS VP1 TO VP4; CORE PROTEINS  
 P2A TO P2C; P3A: GENOME-LINKED PROTEIN VP3; PICORNAIN 3C  
 DE (EC 3.4.22.28) (PROTEASE 3C) (P3C); RNA-DIRECTED RNA POLYMERASE P3D  
 DE (EC 2.7.7.48)].  
 OS Human enterovirus 70 (strain J670/71) (EV 70).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
 OC Enterovirus.  
 OX NCBI\_Taxid=31915;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91037960; PubMed=2172447;  
 RA Ryan M.D., Jenkins O., Hughes P.J., Brown A., Knowles N.J., Booth D.,  
 RA Minor P.D., Almond J.W.;  
 RT "The complete nucleotide sequence of enterovirus type 70:  
 RT relationships with other members of the picornaviridae.";  
 RT J. Gen. Virol. 71:2291-2299(1990).  
 CC -1- FUNCTION: P3C POLYPEPTIDE IS A PROTEASE THAT CLEAVES AT CERTAIN  
 CC Q/G SITES IN THE POLYPEPTIDE. IT MAY BE A CYSTEINE PROTEASE.  
 CC -1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAEDRAL UNITS,  
 CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,  
 CC VP3, AND VP4.  
 CC -1- PTM: SPECIFIC ENZYMOLOGIC CLEAVAGES IN VIVO YIELD NATURE PROTEINS.  
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: D00820; BAA18891.1; -  
 CC PIR: A36253; GNNY7.  
 CC HSSP: P03299; IPOV.  
 CC MEROPS: C03.001; -  
 CC MEROPS: C03.020; -  
 CC InterPro: IPR0001199; Cys-protease-3C.  
 CC InterPro: IPR0003138; Pico\_P1A.  
 CC InterPro: IPR000081; Pico\_P2B.  
 CC InterPro: IPR002527; Pico\_P2B.  
 CC InterPro: IPR000605; RNA\_helicase.  
 CC InterPro: IPR001205; RNA\_pol\_P3D.  
 CC InterPro: IPR001676; RV.  
 CC Pfam: PF00548; Cys-protease-3C; 1.  
 CC Pfam: PF02226; Pico\_P1A; 1.  
 CC Pfam: PF00947; Pico\_P2A; 1.  
 CC Pfam: PF01552; Pico\_P2B; 1.  
 CC Pfam: PF00073; Inv; 3.  
 CC Pfam: PF00680; RNA\_dep\_RNA\_pol; 1.  
 CC Pfam: PF00910; RNA\_helicase; 1.  
 CC ProDom: PD001125; Cys-protease-3C; 1.

DR ProDom: PD001274; Pico\_P2B; 1.  
 DR ProDom: PD001306; Pico\_P2A; 1.  
 DR Polyprotein; Coat protein; Core protein; Transferase; Myristate.  
 KW RNA-directed RNA polymerase; Hydrolase; Thiol protease; Myristate.  
 FT CHAIN 2  
 FT CHAIN 69  
 FT CHAIN 70  
 FT CHAIN 319  
 FT CHAIN 561  
 FT CHAIN 562  
 FT CHAIN 871  
 FT CHAIN 872  
 FT CHAIN 1015  
 FT CHAIN 1113  
 FT CHAIN 1114  
 FT CHAIN 1443  
 FT CHAIN 1532  
 FT CHAIN 1533  
 FT CHAIN 1554  
 FT CHAIN 1555  
 FT CHAIN 1737  
 FT CHAIN 1738  
 FT CHAIN 2194  
 FT LIPID 2  
 FT ACT\_SITE 1701 1701  
 FT ACT\_SITE 1715 1715  
 SQ SEQUENCE 2194 AA; 244590 MW; 15DBAE96EED0673C CRC64;

Query Match 45.8%; Score 22; DB 1; Length 2194;  
 Best Local Similarity 33.3%; Pred. No. 4.6e+02;  
 Matches 6; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 2 EKKXLLXXXXXXVYVXNA 19  
 | | | | |  
 Db 1320 EKGILYTSPLIATITNA 1337

RESULT 8  
 Y850\_HAEIN STANDARD: PRT: 144 AA.  
 ID Y850\_HAEIN  
 AC P4A060:  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 20-NOV-1995 (Rel. 40, Last annotation update)  
 DE HYPOTHEICAL PROTEIN HI0850.  
 GN HI0850.  
 OS Haemophilus influenzae.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
 OC Haemophilus.  
 OX NCBI\_Taxid=727;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=RD / KW20 / ATCC 51907;  
 RX MEDLINE=95350630; PubMed=7542800;  
 RA Ketschmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,  
 RA Kervatage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,  
 RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,  
 RA Scott J.D., Shirley R., Liu L.-I., Glodok A., Kelley J.M.,  
 RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,  
 RA Uettersack T.R., Hanna M.C., Nguyen D.T., Saudak D.M., Brandon R.C.,  
 RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghegan N.S.M.,  
 RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,  
 RA Venter J.C.;  
 RT "Whole-genome random sequencing and assembly of Haemophilus  
 RT influenzae Rd.";  
 RT Science 269:496-512(1995).  
 RL -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE  
 CC (POTENTIAL).  
 CC -1- SIMILARITY: TO E. COLI RSEC AND H. INFLUENZAE HI589.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: U32766; AAC22507.1; -  
 CC TIGR: HI0850; -





EMBL; X07441; CAA30322.1; ALT\_SEQ  
PIR; S01956; S01956.

DR MEROPS: C21.001; Viral\_helcasel.  
 DR InterPro: IPR000606; Viral\_helcasel.  
 DR Pfam: PF01443; Viral\_helcasel.  
 KM Transferase: RNA-directed RNA polymerase; ATP-binding.  
 FT NP\_BIND 976 983 ATP (BY SIMILARITY).  
 FT DOMAIN 1631 1665 POLYMERASE SITE (BY SIMILARITY).  
 SQ SEQUENCE 1844 AA; 206640 MW; A016D58C83D128C CRC64;

Query Match 43.8%; Score 21; DB 1; Length 1844;  
 Best Local Similarity 29.4%; Pred. No. 7.6e+02;  
 Matches 5; Conservative 1; Mismatches 11; Indels 0; Gaps 0;

OY 3 EKKXLLXXXXXXVXNA 19  
 Db 1466 EYAOLESKTOSTIVANA 1482

RESULT 14  
 POLR\_TYMWV STANDARD; PRT; 1844 AA.

CC 01-FEB-1991 (Rel. 17, Created)  
 DT 01-FEB-1991 (Rel. 17, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE RNA REPLICASE POLYPROTEIN (EC 2.7.7.48).  
 OS Turnip yellow mosaic virus (Australian isolate).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Tymovirus.  
 ON NCBI\_TaxID=12155;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=90021184; Pubmed=2800335;  
 RA Keese P., Mackenzie A., Gibbs A.;  
 RT Nucleotide sequence of the genome of an Australian isolate of turnip  
 RT yellow mosaic tymovirus.  
 RL Virology 172:536-546(1989).  
 CC -1- MISCELLANEOUS: THE 206 KDA PROTEIN IS POTENTIALLY A POLYPROTEIN  
 CC (BY HOMOLOGY TO THE LONGER PROTEIN OF THE SINDBIS VIRUS).  
 CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sid.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL: J04373; AAA6592.1; -  
 DR PIR: JQ0109; RMPYM.  
 DR InterPro: IPR000606; Viral\_helcasel.  
 DR Pfam: PF01443; Viral\_helcasel.  
 KM Transferase: RNA-directed RNA polymerase; ATP-binding.  
 FT NP\_BIND 976 983 ATP (BY SIMILARITY).  
 FT DOMAIN 1631 1665 POLYMERASE SITE (BY SIMILARITY).  
 SQ SEQUENCE 1844 AA; 206510 MW; CB447EF05F199A18 CRC64;

Query Match 43.8%; Score 21; DB 1; Length 1844;  
 Best Local Similarity 29.4%; Pred. No. 7.6e+02;  
 Matches 5; Conservative 1; Mismatches 11; Indels 0; Gaps 0;

OY 3 EKKXLLXXXXXXVXNA 19  
 Db 1466 EYAOLESKTOSTIVANA 1482

RESULT 15  
 POLR\_TYMWV STANDARD; PRT; 1844 AA.  
 AC P28477;  
 DT 01-DEC-1992 (Rel. 24, Created)  
 DT 01-DEC-1992 (Rel. 24, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)

DE RNA REPLICASE POLYPROTEIN (EC 2.7.7.48).  
 OS Turnip yellow mosaic virus (isolate TYMC).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Tymovirus.  
 ON NCBI\_TaxID=31751;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92119261; Pubmed=1731998;  
 RA Dreher T.W., Bransom K.L.;  
 RT "Genomic RNA sequence of turnip yellow mosaic virus isolate TYMC, a  
 RT CDNA-based clone with verified infectivity."  
 RL Plant Mol. Biol. 18:403-406(1992).  
 CC -1- MISCELLANEOUS: THE 206 KDA PROTEIN IS POTENTIALLY A POLYPROTEIN  
 CC (BY HOMOLOGY TO THE LONGER PROTEIN OF THE SINDBIS VIRUS).  
 CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL: X16378; CAA34415.1; -  
 DR PIR: S19152; S19152.  
 DR InterPro: IPR000606; Viral\_helcasel.  
 DR Pfam: PF01443; Viral\_helcasel.  
 KM Transferase: RNA-directed RNA polymerase; ATP-binding.  
 FT NP\_BIND 976 983 ATP (BY SIMILARITY).  
 FT DOMAIN 1631 1665 POLYMERASE SITE (BY SIMILARITY).  
 SQ SEQUENCE 1844 AA; 206612 MW; 02CB928FCCCA5EA1 CRC64;

Query Match 43.8%; Score 21; DB 1; Length 1844;  
 Best Local Similarity 29.4%; Pred. No. 7.6e+02;  
 Matches 5; Conservative 1; Mismatches 11; Indels 0; Gaps 0;

OY 3 EKKXLLXXXXXXVXNA 19  
 Db 1466 EYAOLESKTOSTIVANA 1482

Search completed: February 28, 2002, 11:39:52  
 Job time: 308 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 28, 2002, 11:33:59 ; Search time 35.84 Seconds

(without alignments)  
77.544 Million cell updates/sec

Title: US-09-359-426c-1

Sequence: 1 XEKKXXLXXXXXXXXVXNA 19

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 08

Maximum Match 1008

Listing first 45 summaries

Database :

1: SP-archaea:\*  
2: SP-bacteria:\*  
3: SP-fungi:\*  
4: SP-human:\*  
5: SP-invertebrate:\*  
6: SP-mammal:\*  
7: SP-mhc:\*  
8: SP-organellar:\*  
9: SP-phage:\*  
10: SP-plant:\*  
11: SP-rodent:\*  
12: SP-virus:\*  
13: SP-vertebrate:\*  
14: SP-unclassified:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	26	54.2	514	10	09L283	09L283 arbidopsis
2	25	52.1	135	6	09T52	09T52 bos taurus
3	25	52.1	404	6	09M25	09M25 sus scrofa
4	25	52.1	702	4	09H47	09H47 homo sapien
5	25	52.1	703	4	09H521	09H521 homo sapien
6	24	50.0	214	2	007609	007609 bacillus su
7	24	50.0	305	7	098261	098261 homo sapien
8	24	50.0	408	5	044776	044776 caenorhabdi
9	24	50.0	408	13	09PTK6	09PTK6 xenopus lae
10	24	50.0	455	4	09UTR0	09UTR0 homo sapien
11	23	47.9	113	2	09PIS7	09PIS7 campylobact
12	23	47.9	205	2	09IAX7	09IAX7 pseudomonas
13	23	47.9	251	2	09CIC1	09CIC1 laccococcus
14	23	47.9	251	9	09AZO4	09AZO4 bacterioph
15	23	47.9	892	4	09H86	09H86 homo sapien
16	23	47.9	1111	10	09QX8	09QX8 arbidopsis
17	23	47.9	1210	4	09U011	09U011 homo sapien
18	23	47.9	4641	4	075592	075592 homo sapien
19	22	45.8	202	2	068721	068721 yersinia pe

20	22	45.8	253	2	09EZM4	09EZM4 shigella fl
21	22	45.8	266	2	096648	096648 bacillus su
22	22	45.8	323	2	066169	066169 agrobacteri
23	22	45.8	387	10	09FX81	09FX81 arbidopsin
24	22	45.8	437	4	09BSC3	09BSC3 homo sapien
25	22	45.8	653	2	09CG09	09CG09 laccococcus
26	22	45.8	742	5	09TZ14	09TZ14 caenorhabdi
27	22	45.8	937	5	09VKT1	09VKT1 drosophila
28	22	45.8	1114	5	09NJA1	09NJA1 drosophila
29	22	45.8	3085	12	070710	070710 infectious
30	21	43.8	174	13	09W6F2	09W6F2 gallus gall
31	21	43.8	238	1	058713	058713 pyrococcus
32	21	43.8	274	5	09V872	09V872 drosophila
33	21	43.8	304	2	09CJ71	09CJ71 laccococcus
34	21	43.8	391	5	09GVL9	09GVL9 caenorhabdi
35	21	43.8	405	2	084858	084858 chlamydia t
36	21	43.8	413	10	023118	023118 arbidopsin
37	21	43.8	413	10	065688	065688 arbidopsin
38	21	43.8	438	5	09NAN1	09NAN1 caenorhabdi
39	21	43.8	456	5	09XUM7	09XUM7 caenorhabdi
40	21	43.8	457	2	09EXY7	09EXY7 klebsiella
41	21	43.8	458	5	076672	076672 caenorhabdi
42	21	43.8	515	2	09KTM2	09KTM2 vibrio chol
43	21	43.8	534	2	09KTM2	09KTM2 vibrio chol
44	21	43.8	572	5	09U0N9	09U0N9 plasmodium
45	21	43.8	590	2	09RNN9	09RNN9 zymomonas m

## ALIGNMENTS

RESULT	1	PRELIMINARY:	PRT:	514 AA.
ID	09L283			
AC	09L283			
DT	01-OCT-2000 (TREMBlrel. 15, Created)			
DT	01-OCT-2000 (TREMBlrel. 15, Last sequence update)			
DT	01-JUN-2001 (TREMBlrel. 17, Last annotation update)			
DE	HYPOTHETICAL 56.7 KDA PROTEIN.			
GN	T32M21_20.			
OS	Arbidopsis thaliana (Mouse-ear cress).			
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;			
OC	eurosid II; Brassicales; Brassicaceae; Arabidopsids.			
OX	NCBI_TaxID=3702;			
LN	11			
RP	SEQUENCE FROM N.A.			
RA	Beyan M., Terryn N., Ardiles W., Buysnaert C., Dasseville R.,			
RA	De Clerck R., De Keyser A., Neyt P., Rouze P., Van Den Daele H.,			
RA	Villarroel R., Giesen J., Van Montagu M., Bancroft I., Mewes H.W.,			
RA	Rudd S., Lemcke K., Mayer K.F.X.;			
LN	12			
RP	Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.			
RA	SEQUENCE FROM N.A.			
RU	EU Arabidopsis sequencing project;			
RU	Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; A1162875; CAB85548.1;			
DR	InterPro; IPR001798; Kelch.			
DR	Pfam; PF01344; Kelch; 3.			
KW	Hypothetical protein			
SO	SEQUENCE 514 AA; 56659 MW; D362EB63687B984D CRC64;			

Query Match 54.2%; Score 26; DB 10; Length 514;  
Best Local Similarity 38.9%; Pred. No. 28;  
Matches 7; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

OY 2 EEKXXLXXXXXXXXVXNA 19  
DB 407 EEKRALESIAETOVENA 424  
RESULT 2

```

09T552          PRELIMINARY:      PRT:      135 AA.
ID 09H552
AC 09H552:
DT 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
DT 01-JUN-2000 (TREMblrel. 14, Last annotation update)
DE ADIPOCYTE DIFFERENTIATION-RELATED PROTEIN (FRAGMENT).
OS Bos taurus (Bovine)
OC Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Eukaryota; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE.
RX MEDLINE=97157014; PubMed=9003395;
RA Held H.W., Schnolzer M., Keenan T.W.,
RT "Adipocyte differentiation-related protein is secreted into milk as a
  constituent of milk lipid globule membrane.";
  Biochem. J. 320:1025-1030(1996).
ET NON_TER 1
ET NON_CONS 15
ET NON_CONS 16
ET NON_CONS 26
ET NON_CONS 27
ET NON_CONS 33
ET NON_CONS 34
ET NON_CONS 53
ET NON_CONS 83
ET NON_CONS 84
ET NON_CONS 99
ET NON_CONS 100
ET NON_CONS 107
ET NON_CONS 117
ET NON_CONS 116
ET NON_TER 135
SQ SEQUENCE 135 AA; BAB3E9C0D78C589 CRC64;

Query Match          52.1%; Score 25; DB 6; Length 135;
Best Local Similarity 38.9%; Pred. No. 14;
Matches 7; Conservative 1; Mismatches 10; Indels 0; Gaps 0;

OY 2 EEKXXLXXXXXXVXNA 19
   ||| : |||
Db 35 EEKLPILNPTNQVAVNA 52

RESULT 3
ID 09MZE5          PRELIMINARY:      PRT:      404 AA.
AC 09MZE5:
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMblrel. 15, Last annotation update)
DE ADIPOSE DIFFERENTIATION-RELATED PROTEIN (FRAGMENT).
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RX Kim T.H., Yoon D.H., Kim N.S., Jang Y.S., Cheong I.C., Han J.Y.;
RT "Isolation and characterization of a partial cDNA coding for an
  adipose differentiation-related protein.";
  Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
RL EMBL: AF234676; AAF76320.1; -.
DR NON_TER 404
DR NON_CONS 404
SQ SEQUENCE 404 AA; 44491 MW; 6764A09E262CD673 CRC64;

Query Match          52.1%; Score 25; DB 6; Length 404;
Best Local Similarity 38.9%; Pred. No. 44;
Matches 7; Conservative 1; Mismatches 10; Indels 0; Gaps 0;

OY 2 EEKXXLXXXXXXVXNA 19
   ||| : |||
Db 91 EEKLPILNPTNQVAVNA 108

```

```

RESULT 4
ID 09H4H7          PRELIMINARY:      PRT:      702 AA.
AC 09H4H7:
DT 01-MAR-2001 (TREMblrel. 16, Created)
DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE D616B8.1 (SIMILAR TO PRE-MRNA SPLICING RNA HELICASE) (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX Laid G.;
RT Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
CC -1 SIMILARITY: TO HELICASE C-TERMINAL DOMAIN.
CC EMBL: AL023803; CAC03449.3; -.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR001650; Helicase_C.
DR Pfam: PF00271; Helicase_C; 1.
DR SMART: SM00487; DEXDC; 1.
DR SMART: SM00490; HELIC; 1.
KW ATP-binding; Helicase.
FT NON_TER 1
SQ SEQUENCE 702 AA; 78927 MW; 7A2E8DE2F7197746 CRC64;

Query Match          52.1%; Score 25; DB 4; Length 702;
Best Local Similarity 35.3%; Pred. No. 79;
Matches 6; Conservative 1; Mismatches 10; Indels 0; Gaps 0;

OY 2 EEKXXLXXXXXXVXN 18
   ||| : |||
Db 24 EEROSTAENSGTIVYN 40

RESULT 5
ID 09H5Z1          PRELIMINARY:      PRT:      703 AA.
AC 09H5Z1:
DT 01-MAR-2001 (TREMblrel. 16, Created)
DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE CDNA: FLJ22759 FIS, CLONE KAI1A0875.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE-LEAL MUCOSA;
RC Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T., Matsumura K.,
RA Nakajima Y., Mizuno T., Morinaga M., Tanigami A., Fujiwara T., Ono T.,
RA Yamada K., Fujii Y., Ozaki K., Hiro M., Omori Y., Ota T., Suzuki Y.,
RA Odayashi M., Nishi T., Shibahara T., Tanaka T., Nakamura Y.,
RA Isogai T., Sugano S.;
RT "NEBO human cDNA sequencing project.";
  Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
RL -1 SIMILARITY: TO HELICASE C-TERMINAL DOMAIN.
CC EMBL: AK026412; BAB15476.1; -.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR001650; Helicase_C.
DR Pfam: PF00271; Helicase_C; 1.
DR SMART: SM00487; DEXDC; 1.
DR SMART: SM00490; HELIC; 1.
KW ATP-binding; Helicase.
FT NON_TER 703
SQ SEQUENCE 703 AA; 78811 MW; CFC9AB3F90CE580C CRC64;

Query Match          52.1%; Score 25; DB 4; Length 703;

```

Best Local Similarity 35.3%; Pred. No. 79;  
Matches 6; Conservative 1; Mismatches 10; Indels 0; Gaps 0;  
Db 24 EEROSLANSTCTIVVN 40

RESULT 6  
ID 007609 PRELIMINARY; PRT: 214 AA.  
AC 007609;  
DT 01-JUL-1997 (TREMBlrel. 04, Created)  
DT 01-JUL-1997 (TREMBlrel. 04, Last sequence update)  
DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)  
DE HYPOTHETICAL 22.8 KDA PROTEIN.  
GN YHK.  
OS Bacillus subtilis.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
OC Bacillus/Staphylococcus group; Bacillus.  
OX NCBI\_TaxID=1423;  
[1]  
SEQUENCE FROM N.A.  
STRAIN=168:  
Noback M.A., Terpstra P., Holsappel S., Venema G., Bron S.;  
Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.  
[2]  
SEQUENCE FROM N.A.  
STRAIN=168:  
MEDLINE=98044033; PubMed=9384377;  
Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,  
Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,  
Bortis R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,  
Brouillet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,  
Choi S.K., Codani J.J., Conerton I.F., Cummings N.J., Daniel R.A.,  
Dentrot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,  
Ertlan K.D., Errington J., Fabbre C., Ferrati E., Foulger D.,  
Fritz C., Fujita M., Fujita Y., Funa S., Gallizzi A., Galleron N.,  
Gilm S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,  
Gulseppl G., Guy B.J., Haga K., Haelech J., Harwood C.R., Henaut A.,  
Hilbert H., Holsappel S., Hosono S., Hullo M.F., Iaya M., Jones L.,  
Jorle B., Karamata D., Kasahara Y., Kjaer-Bianhard M., Klein C.,  
Kobayashi Y., Koelter P., Koningsstein G., Krogh S., Kumano M.,  
Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,  
Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medigue C.,  
Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,  
Noone D., O'Reilly M., Ogawa K., Ogihara A., Oudega B., Park S.H.,  
Parro V., Pohl T.M., Portetelle D., Portollik S., Prescott A.M.,  
Prescan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,  
Rieger M., Rivoita C., Rocha E., Roche B., Rose M., Sadale Y.,  
Sato T., Scanlan E., Schleich S., Schroeter R., Scofield F.,  
Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,  
Sorokin A., Tacccon E., Takagi T., Takahashi H., Takemaru K.,  
Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,  
Tosato V., Uchlyama S., Vandenbol M., Vannier F., Vasseroiti A.,  
Winters P., Wipat R., Wiedler E., Wiedler H., Wietzenegger T.,  
Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa K., Yata K.,  
Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa K., Yata K.,  
"The complete genome sequence of the gram-positive bacterium Bacillus  
subtilis";  
Nature 390:249-256(1997).  
[3]  
SEQUENCE FROM N.A.  
STRAIN=168:  
Kunst F., Ogasawara N., Yoshikawa H., Danchin A.;  
Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL: Y14083; CA74532.1;  
DR EMBL: Z99109; CAB12866.1;  
KW Hypothetical protein; Complete proteome.  
SEQUENCE 214 AA; 22760 MW; 7CBB20BC5EE2F4B1 CRC64;

Query Match 50.0%; Score 24; DB 2; Length 214;

Best Local Similarity 29.4%; Pred. No. 45;  
Matches 5; Conservative 2; Mismatches 10; Indels 0; Gaps 0;  
Db 36 EOKASLEAGAGAVLAN 52

RESULT 7  
ID 098261 PRELIMINARY; PRT: 305 AA.  
AC 098261;  
DT 01-MAY-1999 (TREMBlrel. 10, Created)  
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)  
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
DE BTL-II (MHC CLASS II ASSOCIATED BUTYROPOLIN-LIKE) (FRAGMENT).  
GN BTL-II.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
[1]  
SEQUENCE FROM N.A.  
Phillips S.;  
Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.  
CC -1- SIMILARITY: TO IMMUNOGLOBULIN AND MAJOR HISTOCOMPATIBILITY COMPLEX  
DOMAIN  
DR EMBL: AL034394; CAB8473.1; -;  
DR InterPro: IPR003599; 19;  
DR InterPro: IPR003600; 19\_1like.  
DR InterPro: IPR003006; 19\_MHC.  
DR Pfam: PF00047; 19; 2.  
DR SMART: SM00409; 19; 1.  
DR SMART: SM00410; 19\_1like; 1.  
KW MHC.  
FT NON\_TER  
SEQUENCE 305 AA; 33497 MW; 8207B980E7B826D2 CRC64;  
SO  
Query Match 50.0%; Score 24; DB 7; Length 305;  
Best Local Similarity 33.3%; Pred. No. 65;  
Matches 6; Conservative 2; Mismatches 10; Indels 0; Gaps 0;  
Db 51 QDKDGLFYVATLVVNA 68

RESULT 8  
ID 044776 PRELIMINARY; PRT: 408 AA.  
AC 044776;  
DT 01-JUN-1998 (TREMBlrel. 06, Created)  
DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)  
DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)  
DE F33D11.2 PROTEIN.  
GN F33D11.2.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
[1]  
SEQUENCE FROM N.A.  
STRAIN=BRISTOL N2;  
MEDLINE=94150718; PubMed=7906398;  
Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,  
Bonfield J., Burton J., Connell M., Copey T., Cooper J., Coulson A.,  
Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,  
Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,  
Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,  
Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,  
Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Shownkeen R.,  
Smaildon N., Smith A., Sonhammer E., Staden R., Sulston J.,

RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,  
 RA Watson A., Weinstock L., Wilkinson-Sprat J., Woldman P., III of C.  
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
 RT elegans.";  
 RL Nature 368:32-38(1994).  
 RN [12]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RA Sammons L., Woldmann P., Mullen G.;  
 RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.  
 RN [13]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RA Waterston R.;  
 RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF039720; AAB9696.1;  
 SQ SEQUENCE 408 AA; 45741 MW; DB10968F6FCED2 CRC64;

Query Match 50.0%; Score 24; DB 5; Length 408;  
 Best Local Similarity 33.3%; Pred. No. 89;  
 Matches 6; Conservative 1; Mismatches 11; Indels 0; Gaps 0;

OY 2 EEKXXLXXXXXXXXXXVNA 19  
 ||| : |||  
 DB 311 EKKKSTIEKTAVALNA 328

RESULT 9  
 O9PTK6 PRELIMINARY; PRT; 408 AA.  
 AC O9PTK6;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)  
 DE FATVG.  
 GN FATVG.  
 OS Xenopus laevis (African clawed frog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;  
 OC Xenopodidae; Xenopus.  
 OX NCBI\_TaxID=8335;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-20002427; PubMed-10529413;  
 RL "fatvg encodes a new localized RNA that uses a 25-nucleotide element (FVTEI) to localize to the vegetal cortex of Xenopus oocytes.";  
 DR EMBL: AF184090; AAF19611.1;  
 SQ SEQUENCE 408 AA; 44672 MW; BAC91B0C9B796C29 CRC64;

Query Match 50.0%; Score 24; DB 13; Length 408;  
 Best Local Similarity 33.3%; Pred. No. 89;  
 Matches 6; Conservative 2; Mismatches 10; Indels 0; Gaps 0;

OY 2 EEKXXLXXXXXXXXXXVNA 19  
 ||| : |||  
 DB 93 EKKPLIYOPSDIVSNA 110

RESULT 10  
 O9UIRO PRELIMINARY; PRT; 455 AA.  
 AC O9UIRO;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE BUTYROPHILIN-LIKE;  
 GN BUTI.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Stammers M., Rowen L., Rhodes D., Trowsdale J., Beck S.;  
 RT "Characterization of BTL-11: A polymorphic butyrophilin-like gene  
 RT located at the border of the MHC class II and class III regions in  
 RT human and mouse.";  
 RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: TO IMMUNOGLOBULIN AND MAJOR HISTOCOMPATIBILITY COMPLEX  
 CC DOMAIN.  
 DR EMBL: AF186593; AAF05530.1; JOINED.  
 DR EMBL: AF186588; AAF05530.1; JOINED.  
 DR EMBL: AF186589; AAF05530.1; JOINED.  
 DR EMBL: AF186590; AAF05530.1; JOINED.  
 DR EMBL: AF186591; AAF05530.1; JOINED.  
 DR EMBL: AF186592; AAF05530.1; JOINED.  
 DR InterPro: IPR003599; 19.  
 DR InterPro: IPR003600; 19\_1like.  
 DR InterPro: IPR003006; 19\_MHC.  
 DR Pfam: PF00047; 19; 3.  
 DR SMART: SM00409; IG; 2.  
 DR SMART: SM00410; IG\_Like; 1.  
 SQ SEQUENCE 455 AA; 50436 MW; 4194025C7416F839 CRC64;

Query Match 50.0%; Score 24; DB 4; Length 455;  
 Best Local Similarity 33.3%; Pred. No. 1e+02;  
 Matches 6; Conservative 2; Mismatches 10; Indels 0; Gaps 0;

OY 2 EEKXXLXXXXXXXXXXVNA 19  
 ||| : |||  
 DB 194 QDKDLFYEAETLVVNA 211

RESULT 11  
 O9PIST PRELIMINARY; PRT; 113 AA.  
 AC O9PIST;  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
 DE HYPOTHETICAL PROTEIN CU0202C.  
 GN CU0202C.  
 OS Campylobacter jejuni.  
 OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;  
 OC Campylobacter.  
 OX NCBI\_TaxID=197;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-NCTC 11168;  
 RX MEDLINE-20150912; PubMed-10688204;  
 RA Parkhill J., Wren B.W., Mungall K., Kettle J.M., Churcher C.,  
 RA Basham D., Chillingworth T., Davies R.M., Feltham T., Holtroyd S.,  
 RA Jagels K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W.,  
 RA Quail M.A., Rajandream M.A., Rutherford K.M., Van Vliet A.H.M.,  
 RA Whitehead S., Barrett B.G.;  
 RT "The genome sequence of the food-borne pathogen Campylobacter jejuni  
 RT reveals hypervariable sequences.";  
 RL Nature 403:665-668(2000).  
 DR EMBL: AL139074; CAB72685.1;  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 113 AA; 12679 MW; EB8C40AD614357B0 CRC64;

Query Match 47.9%; Score 23; DB 2; Length 113;  
 Best Local Similarity 33.3%; Pred. No. 45;  
 Matches 6; Conservative 1; Mismatches 11; Indels 0; Gaps 0;

OY 2 EEKXXLXXXXXXXXXXVNA 19  
 ||| : |||  
 DB 40 EKKTFVSDKITGVVNA 57



RESULT 12  
 ID 0914X7 PRELIMINARY: PRT: 205 AA.  
 AC 0914X7:  
 DT 01-MAR-2001 (TREMBlrel. 16, Created)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
 DE PROBABLE FIMBRIAL SUBUNIT PROTEIN.  
 GN PA0992.  
 OS Pseudomonas aeruginosa.  
 OC Bacteria: Proteobacteria; gamma subdivision; Pseudomonadaceae;  
 OC Pseudomonas.  
 CX NCBI\_TaxID=287;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=PA01;  
 RX MEDLINE=20437337; PubMed=10984043;  
 RA Hickey M.J., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,  
 Stover C.K., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,  
 Garber R.L., Goltzy L., Tolentino E., Westbrock-Wadman S., Yuan Y.,  
 Brody L.L., Coulter S.N., Folger K.R., Kas A., Landig K., Lim R.M.,  
 Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,  
 Ralston J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.,  
 "Complete genome sequence of Pseudomonas aeruginosa PA01, an  
 opportunistic pathogen."  
 RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an  
 opportunistic pathogen."  
 RL Nature 406:959-964(2000).  
 DR EMBL: AE004532; AAG04381.1;  
 DR InterPro: IPR000259; Fimbril.  
 DR Pfam: PF00419; Fimbril; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 205 AA; 20046 MW; 40B34453DB19E301 CRC64;

Query Match 47.9%; Score 23; DB 2; Length 205;  
 Best Local Similarity 37.5%; Pred. No. 85;  
 Matches 6; Conservative 0; Mismatches 10; Indels 0; Gaps 0;  
 QY 4 KXXLXXXXXXVXNA 19  
 DB 6 KSTLAANVASLVYVNA 21

RESULT 13  
 ID 09C1C1 PRELIMINARY: PRT: 251 AA.  
 AC 09C1C1:  
 DT 01-JUN-2001 (TREMBlrel. 17, Created)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
 DE PROPHAGE P11 PROTEIN 08.  
 GN P1108.  
 OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).  
 OC Bacteria: Firmicutes: Bacillus/Clostridium group; Streptococcaceae;  
 OC Lactococcus.  
 CX NCBI\_TaxID=1360;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=IL1403;  
 RA Bolotin A., Mincer P., Manger S., Tallon O., Malarme K.,  
 Weissenbach J., Ehrlich S.D., Sorokin A.,  
 "The complete genome sequence of the lactic acid bacterium Lactococcus  
 lactis."  
 RT Lactis.  
 RL Genome Res. 0:0-0(2001).  
 DR EMBL: AE006281; AAK04541.1;  
 KW Complete proteome.  
 SQ SEQUENCE 251 AA; 28744 MW; A7E842DE042B03A1 CRC64;

Query Match 47.9%; Score 23; DB 2; Length 251;  
 Best Local Similarity 35.3%; Pred. No. 1.1e+02;  
 Matches 6; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 2 EEKXXLXXXXXXVXN 18  
 DB 133 EEKQALRALNSTLAVEN 149  
 RESULT 14  
 ID 09A2Q4 PRELIMINARY: PRT: 251 AA.  
 AC 09A2Q4:  
 DT 01-JUN-2001 (TREMBlrel. 17, Created)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
 DE ANTI-REPRESSOR.  
 GN ORF8.  
 OS bacteriophage b1309.  
 OC Viruses: dsDNA viruses, no RNA stage; Tailed phages; siphoviridae.  
 CX NCBI\_TaxID=151537;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=2111349; PubMed=11160885;  
 RA Chopin A., Bolotin A., Sorokin A., Ehrlich S.D., Chopin M.-C.,  
 "Analysis of six prophages in Lactococcus lactis IL1403: different  
 genetic structure of temperate and virulent phage populations."  
 RT Nucleic Acids Res. 29:644-651(2001).  
 RL [2]  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Chopin A., Bolotin A., Sorokin A., Ehrlich S.D., Chopin M.-C.,  
 Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF23670; AAK08356.1;  
 SQ SEQUENCE 251 AA; 28744 MW; A7E842DE042B03A1 CRC64;

Query Match 47.9%; Score 23; DB 9; Length 251;  
 Best Local Similarity 35.3%; Pred. No. 1.1e+02;  
 Matches 6; Conservative 0; Mismatches 11; Indels 0; Gaps 0;  
 QY 2 EEKXXLXXXXXXVXN 18  
 DB 133 EEKQALRALNSTLAVEN 149

RESULT 15  
 ID 09H8E6 PRELIMINARY: PRT: 892 AA.  
 AC 09H8E6:  
 DT 01-MAR-2001 (TREMBlrel. 16, Created)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
 DE CDNA FLJ13704 F15, CLONE PLACE2000274, WEAKLY SIMILAR TO DYNEIN BETA  
 DE CHAIN, CILIARY.  
 OS Homo sapiens (Human).  
 OC Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi;  
 OC Mammalia: Eutheria: Primates: Catarrhini: Hominiidae: Homo.  
 CX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=PLACENTA;  
 RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,  
 Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,  
 Tanase T., Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K.,  
 Arita M., Nabekura T., Ishii S., Kawai Y., Saito K., Yamamoto J.,  
 Wakamatsu A., Nakamura Y., Nagahara K., Masuno Y., Oshima A.,  
 "WEDD human cDNA sequencing project."  
 RT Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
 RL EMBL: AK023766; BAB14671.1;  
 DR InterPro: IPR001230; Preynlin.  
 DR PROSITE: PS00294; PRENYLATION;  
 SQ SEQUENCE 892 AA; 102234 MW; 155C299D6944854D CRC64;

Query Match 47.9%; Score 23; DB 4; Length 892;  
 Best Local Similarity 27.8%; Pred. No. 4.1e+02;  
 Matches 5; Conservative 3; Mismatches 10; Indels 0; Gaps 0;

Fri Mar 1 07:28:38 2002

us-09-359-426c-1.rspt

Page 6

OY 2 EEXXXLXXXXXXVXNA 19  
| | : |  
DB 118 EEPDLEAKNOLITISNA 135

Search completed: February 28, 2002, 11:39:32  
Job time: 333 sec